BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 9/127, 9/16, 9/51

A1

(11) International Publication Number: WO 96/13250

(43) International Publication Date: 9 May 1996 (09.05.96)

(21) International Application Number: PCT/US95/13749

(22) International Filing Date: 20 October 1995 (20.10.95)

(30) Priority Data:
08/331,393 27 October 1994 (27.10.94) US

(71) Applicant: AMGEM INC. [US/US]; Amgen Center, 1840 Dehavilland Drive, Thousand Oaks, A 91320-1789 (US).

(72) Inventors: HABBERFIELD, Alan, D.; 17838 Castellammare Drive, Pacific Palisades, CA 90272 (US). JENSEN-PIPPO, Kathleen; 4343 Cedardale Road, Moorpark, CA 93021 (US).

(74) Agents: ODRE, Steven, M. et al.; Amgen Inc., Amgen Center, 1840 Dehavilland Drive, Thousand Oaks, CA 91320-1789 (US).

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: COMPOSITIONS FOR INCREASED BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS

(57) Abstract

The present invention involves compositions and methods for enhancing the bioavailability of therapeutic agents. In particular, the bioavailability of the therapeutic agent is enhanced by combining the agent with an invasion proficient protein, wherein the protein facilities the transport of the therapeutic agent across the gastrointestinal barrier.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

COMPOSITIONS FOR INCREASED BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS

FIELD OF THE INVENTION

The present invention relates to the enhancement of the bioavailability of orally delivered therapeutic agents. In particular, the invention involves improving the bioavailability of therapeutic agents by combining them with a suitable transport promoter which is capable of facilitating the penetration of the therapeutic agent across epithelial and endothelial cell barriers. The transport promoter of the present invention is preferably an invasion proficient bacterial coat protein which, when combined with a therapeutic agent, can effectuate the penetration of the therapeutic agent through the gastrointestinal lining.

BACKGROUND OF THE INVENTION

The common routes of therapeutic agent administration are enteral (oral) and parental (intravenous, subcutaneous, and intramuscular) routes of administration. The intravenous route is advantageous for emergency use when a very rapid and predictable increase in blood level of the therapeutic agent is necessary. In addition, the intravenous route allows for easy dosage adjustments and is useful for administering large volumes of a drug. Intravenous drug administration, however, has several limitations. One problem is the risk of adverse effects resulting from the rapid accumulation of a high concentration of the therapeutic agent in plasma

and/or tissues. Also, repeated injections by the intravenous route may cause discomfort to the patient. In addition, the delivery is inconvenient as often it is administered by a health care provider.

The oral administration of a therapeutic agent is generally more convenient, economical and acceptable. Oral delivery is by far the most popular delivery method where the drug is intended to be absorbed by the gastrointestinal tract. There are, however, several problems associated with the oral delivery of therapeutic agents. For example, oral administration is limited when the therapeutic agent is not efficiently absorbed by the gastrointestinal tract. Unlike the administration of a therapeutic agent by injection, which circumvents the highly protective barriers of the human body, the absorption of a therapeutic agent by the gastrointestinal tract may be inefficient for poorly soluble, slowly absorbed, or unstable therapeutic preparations. As a result, many important therapeutic agents, which are not effectively absorbed when administered orally, are currently delivered by injection.

In particular, the delivery of polypeptide and protein therapeutic agents via the gastrointestinal tract is especially difficult because of the inherent instability of such materials and the poor permeability of the intestinal mucosa to high molecular weight substances. The gastrointestine is an organ of the body that is specifically developed to physically, chemically and enzymatically break down ingested nutrients. The gastrointestine is also responsible for the uptake of nutrients into the body and for the elimination of waste. The gastro-intestinal tract includes the stomach and intestine. The stomach is specifically designed for the digestion of nutrients, the stimulation of other regions of the

gut to secrete, the storage of food, and the release of chyme into the intestine at a controlled rate. Nutrient uptake is not an important function of the stomach. The small intestine includes the duodenum, jejunum and ileum. Distal to the stomach is the duodenum, where neutralization of the acidic chyme occurs. Surfactants for lipid digestion and proteases for protein breakdown are also secreted into the duodenum. There is little absorption in this section of the gut. Uptake of the nutrient breakdown products mainly occurs in the lower small intestine: the jejunum and the ileum are 2.8 meters and 4.2 meters in length respectively, and have a combined surface area of 460 m^2.

The large intestine, which is composed of the cecum and the colon, is responsible for the storage of waste, and also for water and salt balance. There is little enzyme activity in this section of the gut, and it is the least permeable section of the gastrointestinal tract.

The majority of the surface of the small and large intestine is lined by a layer of epithelial cells called the enterocytes, which are specialized villus absorptive cells. The lining of the gut is also composed of a mucus lining which acts as an unstirred water layer (1). The mucus is a barrier to macromolecules with a molecular weight greater than 17 KDa (2). The enterocyte lining forms a tight lipid barrier to peptides having a molecular weight as low as 500 Da (3). Therefore, the lining of the gut is composed of an efficient barrier to both lipophilic and hydrophilic molecules due to the mucus and the enterocyte linings, respectively. The oral administration of a large, macromolecular therapeutic agent is, therefore, very limited by the barrier effect of the gastrointestinal lining. This is

certainly true of the recombinant therapeutic proteins.

The gastrointestinal tract, however, cannot be a complete barrier to all macromolecules because many macromolecules are required for nutrient intake. These include, among others, amino acids, glucose and vitamins. For such molecules, specific transport mechanisms exist. Amino acids and glucose are taken up by transporters situated in the lumenal or apical membrane domains of the enterocytes. Receptors for vitamin uptake are also present in the apical domain of the enterocyte lining.

In addition, certain microorganisms, including both viruses (<100 nm in diameter) and bacteria (>1μm in diameter), are able to invade the body from the gut by crossing the epithelial barrier. Certain cells of the immune system, including neutrophils and macrophages, are also able to permeate both epithelial and endothelial barriers.

Bacteria that invade the enterocyte barrier include, *Yersinia*, *Salmonella*, *Shigella* and *Listeria*. In the case of *Yersinia*, the method of attachment to the cell surface and invasion into the cell has been characterized. In *Yersinia pseudotuberculosis* and in *Yersinia enterocolitica*, a protein termed invasin (INV) is expressed on the surface of the bacteria. It has been shown that the INV protein is able to bind to the β_1 integrin family of receptors (4, 5). The integrin receptor family belongs to a group of molecules termed the adhesion receptors and is involved in promoting cell attachment to the extracellular matrix (6). Following binding of the INV protein to the cell, internalization of the protein occurs (7). This event has been demonstrated in HEp-2 cells, which are epithelial-like cells from the larynx, and in some other epithelial cells. The

invasion event has not been demonstrated in the enterocyte cells.

Another invasion-mediating protein identified in *Yersinia enterocolitica* has been termed the AIL protein (for attachment-invasion-locus) (8). The receptor utilized by this protein is as yet unknown, and as with INV, the binding and invasion event has not been demonstrated for gut epithelium.

In vivo studies have shown that *Yersinia* can invade the body from the gut through the Peyers Patches (9, 10). No studies have shown that the INV and AIL proteins are able to mediate binding and invasion of the enterocytes lining the gut.

The delivery of a therapeutic agent through the enterocyte lining would be preferable, as compared to Peyers Patch uptake, because the latter are known to be variable from species to species and between individuals of the same species. In addition, materials delivered through the Peyers Patch are more effectively delivered as an antigen.

CURRENT METHODS OF DRUG DELIVERY

The efficacy of an orally administered therapeutic agent depends on the agent being absorbed from the gastrointestinal tract into the circulation. The permeability barrier of the gut epithelium is perhaps the most limiting factor to the reproducible oral absorption of therapeutic agents.

One previous attempt to circumvent non-parental bioavailability problems involved intranasal administration of a therapeutic agent. Investigators have also attempted to pass therapeutic agents across the skin through the use of chelating agents, bile salts and surfactants. Similar materials have been used to increase the absorption of therapeutic agents

from the gastrointestinal tract (11). Other investigators have attempted to increase bioavailability from the gastrointestinal tract through the use of liposome-entrapped therapeutic agents.

Liposomes have also been used as a means for target-specific delivery of an encapsulated biologically active material. Liposomes have been attached to materials such as viral membrane proteins, antibodies, streptavidin, transferrin and other ligands as a means of directing the therapeutic agent to the target cell (12). The results of such delivery methods, however, have not demonstrated that the liposome is an effective means for promoting the bioavailability of orally administered proteins. In fact, liposomes alone or attached to such site-specific ligands are unlikely to facilitate absorption of orally delivered agents because liposomes typically are degraded in the lumen of the gut.

Invasive microorganisms have been used to transfer materials into host cells. Isberg et al. (13) describe the genetic transfer of INV or AIL genes into a microorganism to impart an invasive phenotype to that microorganism. The modified microorganism is then used as a vaccine to introduce a pathogen of interest into a host cell. While this technique describes the introduction of exogenous INV and AIL genes to impart an invasive capability on a microorganism, there is no provision for increasing the bioavailability of a therapeutic agent or improving the transport of a therapeutic agent through a mucosal barrier.

Another delivery technique involves nanosphere and microsphere technology (14, 15). This technology is based upon the observed uptake of such microspheres into the body through the M cells of the Peyers

Patches in the gastrointestinal tract. There is, however, no moiety involved that would enhance the uptake of such particles. The delivery of a therapeutic agent through the Peyers Patches is not an efficient way to orally deliver non-vaccine based therapeutics. A material delivered by this route may be presented to the body as an antigen, and this is not a desired attribute for a non-vaccine therapeutic agent.

Another previously available delivery technique involves the use of proteinoid technology (17). Orally administered delivery systems for insulin, heparin and physostigmine include the use of encapsulating spheres which are predominantly less than 10 microns (μm) in diameter and made of artificial polypeptides. The proteinoids are intended to pass through the gastrointestinal mucosa and thereby deliver a therapeutic agent. One very apparent problem with this system is that the protenoids release the drug component under neutral conditions. Because such conditions are found in the gut, especially in the lower small intestine (i.e., ileum), it would be expected that the proteinoids mainly would release the therapeutic agent into the lumen of the gut rather than transport the therapeutic agent across the gastrointestinal lining.

Another drug delivery technique involves receptor-mediated transcytosis, wherein the amino acid sequences of various growth factors are incorporated into the system (i.e., epidermal growth factor and transforming growth factor alpha) (48). Chimeric molecules or fusion peptides are formed by conjugating the growth factor to a desired protein. The proposed chimeric molecules are transcytosed across epithelial cells via an interaction with growth factor receptors. The chimeric molecule system, however, fails to

provide for the protection of the therapeutic against the gut environment. Moreover, this delivery technique would be dependent on a receptor system which is normally present at low levels on the apical or lumenal domain of the enterocyte. The binding and uptake of growth factors from the lumen of the gut is a non-physiological event.

Notwithstanding the above-noted developments in the arts of cell targeting and drug delivery, it is clear that there is a need for novel compositions which enhance the bioavailability of an orally delivered therapeutic agent. It is not sufficient to merely bind the drug to a target cell.

SUMMARY OF THE INVENTION

A major problem associated with the oral delivery of a therapeutic agent is the hostile environment of the gut, especially to protein and peptide therapeutics. Another problem is the impermeability of the mucosal barrier in the gut, especially to large molecular weight materials.

It is an object of the present invention to increase the bioavailability of orally delivered therapeutic agents, particularly polypeptides and proteins, by providing for the improved transport of such therapeutics across the body's epithelial barriers. It is a further object of the present invention to provide a delivery system wherein the delivery means or transport enhancer is not readily subject to degradation in the gut or prone to the early release of the biologically active material.

It is another object of the present invention to provide a transport enhancer which is not subject to

the low residency time of the proteinoids at the mucosal surface.

The present invention is based on the finding that compositions containing INV or AIL invasive proteins are able to cross the cells of the gastrointestinal tract through an internalization and transcytosis event. This was a novel observation and formed the basis of the current invention concerning the delivery of therapeutic agents.

The present invention provides a delivery system, involving a therapeutic agent and an invasion proficient bacterial protein which transports the therapeutic agent across the gastrointestinal membrane barrier, thereby increasing the oral bioavailability of that agent. The system may optionally include a carrier component such as a liposome or polymer-based particle. In an alternate embodiment, the pharmaceutical composition may involve a fusion protein including the therapeutic moiety and an invasion proficient bacterial protein to effect delivery of the composition across the gastrointestinal tract. In yet another embodiment, the therapeutic moiety and invasion proficient protein may be linked by a degradable peptide sequence.

The delivery system of the present invention provides a composition that is stable in the gut, enhances the uptake of the therapeutic moiety and is expected to cross both the enterocytes and the M cells of the Peyers patches. The system provides an increase in bioavailability as well as a clear advantage over existing particle-based systems that are dependent on non-specific uptake through the antigen-presenting M cells. By increasing the bioavailability of intact and active polypeptide and protein therapeutic agents, the present invention also obviates the need for the parenteral administration of

such therapeutic agents which are otherwise degraded in the gut or relatively unable to cross the gastrointestinal barrier.

DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the oligonucleotide and amino acid sequences of invasin (INV) protein (SEQ ID NO:1).

Figure 2 illustrates the oligonucleotide and amino acid sequences of attachment-invasion-locus (AIL) protein (SEQ ID NO:2).

Figure 3 illustrates the oligonucleotide and amino acid sequences of maltose binding protein (MBP) (SEQ ID NO:3).

Figure 4 illustrates the effect of invasin transfection and expression on the binding of *E. coli* to the human enterocyte Caco-2 cell line.

Figure 5 illustrates the effect of invasin transfection and expression on the internalization of *E. coli* into the human enterocyte Caco-2 cell line.

Figure 6 illustrates the effect of AIL-transfection and expression on the binding of *E. coli* to the human enterocyte Caco-2 cell line.

Figure 7 illustrates the effect of AIL-transfection and expression on the internalization of *E. coli* into the human enterocyte Caco-2 cell line.

Figure 8 summarizes a nine hour study showing the effect of both INV- and AIL-transfection and expression on the internalization of *E. coli* into the non-polarized human enterocyte cell line.

Figure 9 illustrates the polarity of receptor distribution in Caco-2 monolayers grown on Transwell-COL inserts. The distribution of the fibronectin, epidermal growth factor (EGF), taurocholic acid (TA)

and intrinsic factor-vitamin B12 complex (IF-VB12) receptors are shown.

Figure 10 illustrates the surface binding of INV- and AIL-transfected *E. coli* to polarized Caco-2 cell monolayers.

Figure 11 illustrates the internalization of INV- and AIL-transfected *E. coli* into polarized Caco-2 cell monolayers.

Figure 12 illustrates the time course of trancytosis of INV- and AIL-transfected *E. coli* across the polarized Caco-2 cell monolayers.

Figure 13 illustrates specificity of the binding of radiolabelled MBP-INV to the non-polarized Caco-2 cell line.

Figure 14 illustrates the amino acid sequence of a fusion protein of invasin and maltose binding protein (SEQ ID NO:4) using the 192 amino acids from the C-terminal end of INV from *Y. pseudotuberculosis*.

Figure 15 illustrates the amino acid sequence of a fusion protein of attachment-invasion-locus protein and maltose binding protein (SEQ ID NO:5).

Figure 16 illustrates the liposome uptake by Caco-2 cells with and without conjugation to MBP-INV.

(Note: All the points shown in the drawings represent the mean ±SEM where n=3)

DETAILED DESCRIPTION OF THE INVENTION

It is known that many bacteria, viruses and cells of the immune system are able to permeate the epithelial and endothelial barriers of the body through the expression of integral or peripheral membrane proteins. Current investigations of bacterial proteins have revealed at least two proteins

that appear to be involved in the invasion of bacteria into the human host. These invasive proteins have been termed invasin (INV) and attachment-invasion-locus (AIL) proteins. Both proteins have been cloned from *Yersinia enterocolitica*, although INV is also known to exist with large homology in *Y. pseudo-tuberculosis*.

The present invention involves the discovery that the INV and AIL proteins may be used to mediate the transport of therapeutic compositions, including large particles (approximately 1 μm), across the polarized human enterocyte, thereby enhancing the penetration or passage of a therapeutic composition across the gastrointestinal barrier. Moreover, it has been determined that such invasion proteins can be removed from their natural bacterial expression system yet retain the ability to bind the human enterocyte.

These findings lead to the development of the present oral delivery system based upon the combination of a therapeutic agent with the INV or AIL protein or derivatives thereof. The bacterial invasion proteins bind to receptors expressed through the apical or luminal domains of the enterocytes or M cells of the Peyers Patches. In this way, INV and AIL act as bioadhesive agents and thereby increase the residence time of the pharmaceutical composition in the gut. This in itself can increase the bioavailability of the therapeutic agent by promoting uptake of the therapeutic agent. It was further determined, however, that INV and AIL also mediate the movement of the composition either paracellularly or transcellularly across the gastrointestinal tract, and thereby facilitate the transport of the therapeutic agent across the mucosal barrier. The bacterial invasion proteins may also be used for increasing drug transport through other non-invasive routes where the

appropriate receptors are expressed. Such routes may include nasal, ocular, rectal, vaginal, pulmonary and transdermal routes of administration.

In one embodiment of the present invention, the bacterial invasion protein is indirectly associated with the therapeutic agent through a linking means such as a polymer chain, or directly associated with the therapeutic agent by a chemical means. An alternative embodiment of the present invention is based upon the incorporation of a therapeutic agent into or onto a carrier that is associated with the bacterial invasion protein, such as INV and AIL or fragments or derivatives thereof. The bacterial invasion protein might be bound to, encapsulated within, incorporated in the structure of, or merely combined with the carrier component. Microparticles and liposomes are exemplary of the carrier component in such a delivery system.

The terms "therapeutic agent", "pharmaceutical", "biologically active material" and "drug" may be used interchangeably, and as used herein, preferably include proteins, hormones and/or medicinal peptides useful for treating a medical or veterinary disorder, preventing a medical or veterinary disorder, or regulating the physiology of a human being or animal. Suitable therapeutic agents include cytokines, as well as a wide range of cytotoxic drugs, muscle relaxants, antihypertensives, analgesics, steroids, vitamins, sedatives and hypnotics, antibiotics, chemotherapeutic agents, prostaglandins and radiopharmaceuticals.

The terms "transport enhancer", "transporting ligand" and "ligand" may be used interchangeably, and as used herein, preferably include bacterial protein molecules which, when conjugated to a therapeutic agent, are capable of increasing the delivery of the

therapeutic agent across a mucosal membrane such as the gastrointestinal barrier. In preferred embodiments, "transport enhancer" is intended to include invasion proficient bacterial coat proteins, or fragments or analogs thereof. Such bacterial invasion proteins may be isolated from bacterial cultures or can be produced by known recombinant or synthetic techniques. Methods of isolating and purifying MBP-INV fusion proteins have previously been described (17, 18), but they have not previously been used in the compositions and methods and of the present invention.

In its basic form, the drug delivery system of the present invention is composed of a transport enhancer and the desired therapeutic agent. In an alternate form, the drug delivery system includes an additional component: a carrier moiety. Thus, the pharmaceutical compositions of the present invention may include a transport enhancer such as a bacterial invasion protein. The transport enhancer is associated with or attached to a carrier component, which in preferred embodiments include latex microspheres or liposomes such as those composed of dipalmitoylphosphatidyl-ethanolamine (DPPC):cholesterol (chol):N-glutaryl-dioleoyl-phosphatidylethanolamine (NG-DOPE). The therapeutic agent can be incorporated into or onto the carrier by various methods known in the art or it may be attached to or associated with the transport enhancer.

Exemplary transport enhancers include invasion proficient bacterial proteins such as INV and AIL. Exemplary amino acid and nucleotide sequences of the INV and AIL proteins are illustrated in Figures 1 and 2, respectively, as well as Sequence ID NOs:1 and 2. INV, an 835 amino acid single chain polypeptide, has

been well characterized in the art (20). AIL, a 162 amino acid single chain polypeptide, has also been well characterized in the art (21).

The receptor binding region of INV involves the 192 amino acids at the C-terminal end of the protein (17). This region has been shown to retain the binding affinity of the bacterial invasion protein, and therefore, any sequence containing this region would be suitable for use in the present invention.

The receptor binding regions of AIL which are necessary or sufficient for binding to the bacterial protein receptor would include all or some of the regions from the four extracellular loops (22). These regions include the following sequences:

Loop 1	QSHVKENGYTLDNDPK
Loop 2	HQGYDFFYGSNKFGHGDVD
Loop 3	HGKVKASVFDESISASKT
Loop 4	KLDSIKVG

Invasion proficient bacterial proteins suitable for use in the present invention may be derived from a variety of DNA sequences encoding such proteins. The selected DNA sequence may be a nucleic acid molecule encoding the invasive protein (e.g., an INV or AIL protein including sequences as set forth in Figures 1 and 2) or their complementary strands, naturally occurring allelic variants, sequences capable of hybridizing to a protein-coding area of such DNA sequences under stringent conditions, and sequences which, but for degeneration, would hybridize with the protein-coding area of these defined DNA sequences.

Suitable invasion proficient bacterial proteins also include derivatives of the amino acid sequences. Such derivatives could consist of a truncated form of the invasive protein, especially with deletion of the sequence from the amino terminal end of the INV protein as described above. Such small molecule derivatives of the bacterial proteins are advantageous in that they are less likely to be immunogenic.

Further modifications in the peptides or DNA sequences encoding the invasion proficient bacterial proteins can be made by one skilled in the art using known techniques. Modifications of interest in the protein sequences may include the replacement, insertion or deletion of a selected amino acid residue. Naturally occurring amino acids may be divided into groups based upon common side chain properties:

Hydrophobic:	norleucine, Met, Ala, Val, Leu, Ile
Neutral hydrophilic:	Cys, Ser, Thr
Acidic:	Asp, Glu
Basic:	Asn, Gln, His, Lys, Arg
Residues that influence chain orientation:	Gly, Pro
Aromatic:	Trp, Tyr, Phe

Nonconservative substitutions will entail exchanging a member of one of these classes for another. Other exemplary substitutions are illustrated in Table 1.

Table 1

Original Residue	Exemplary Substitution	Preferred Substitution
Ala (A)	Ile, Leu, Val	Val
Arg (R)	Asn, Gln, Lys	Lys
Asn (N)	Arg, Gln, His, Lys	Gln
Asp (B)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro	Pro
His (H)	Arg, Asn, Gln, Lys	Arg
Ile (I)	Ala, Leu, Met, Phe, Val, norleucine	Leu
Leu (L)	Ala, Ile, Met, Phe, Val, norleucine	Ile
Lys (K)	Arg, Asn, Gln	Arg
Met (M)	Ile, Leu, Phe	Leu
Phe (F)	Ala, Ile, Leu, Val	Leu
Pro (P)	Gly	Gly
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr	Tyr
Tyr (Y)	Phe, Ser, Thr, Trp	Phe
Val (V)	Ala, Ile, Leu, Met, Phe, norleucine	Leu

Mutagenic techniques for making such replacements, insertions or deletions are well known to those

skilled in the art (23) Conservative changes of 1 to 20 amino acids are preferred. Preferred peptides may be generated by proteolytic or glycolytic enzymes, or by direct chemical synthesis.

The selected bacterial adhesion protein may also be modified to facilitate production and handling of the composition. For example, the appropriate invasion protein or amino acid sequence may be produced to include an additional peptide or protein component, such as the maltose binding protein (MBP), which can enhance the purification of the protein from the recombinant expression system. Figure 3 (SEQ ID NO:3) depicts the amino acid (and nucleotide sequences of the maltose binding protein. Additions or substitutions to the INV and AIL amino acid sequences may also be used to facilitate the attachment or immobilization of the transport enhancer to or on the pharmaceutical agent or carrier component of the pharmaceutical composition, thereby promoting the retention of the transport enhancer. This could include, for example, the addition of a cysteine residue to the N-terminal end of the sequence to facilitate chemical conjugation by disulfide bridging, using for instance maleimide. Other deletions, substitutions or additions to the amino acid sequence may have the effect of stabilizing the transport enhancer in solution or in the gut or in the serum.

Suitable transport enhancers are selected from proteins or polypeptides which demonstrate an appropriate binding affinity for the receptors found in the cells that form the membrane barrier through which the pharmaceutical composition is to be transported. The amino acid sequences of the INV or AIL proteins demonstrate such a binding affinity for the receptors found in the gut. Preferably, the transport enhancer will also have some specificity for

the cell type that is being targeted. The amino acid sequences of the INV or AIL proteins demonstrate such a specificity for human enterocytes, which is advantageous for gastrointestinal delivery.

The novel compositions of the present invention can be combined with conventional pharmaceutically acceptable excipients suitable for the formulation of therapeutic compositions. As used herein, the term "pharmaceutically acceptable excipient" means a non-toxic, inert solid, semi-solid or liquid component included withing the pharmaceutical formulation. Such pharmaceutically acceptable carriers include, but are not limited to, fillers, diluents, encapsulating materials, solvents or formulation agents, involved in facilitating the carrying or delivery of the pharmaceutical agent. Some examples of the materials that can serve as pharmaceutically acceptable excipients include: sugars, such as lactose, glucose and sucrose; starches such as corn starch and potato starch; cellulose and its derivatives such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatin; talc; excipients such as cocoa butter and suppository waxes; oils such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; glycols, such as propylene glycol; polyols such as glycerin, sorbitol, mannitol and polyethylene glycol; esters such as ethyl oleate and ethyl laurate; agar; buffering agents such as magnesium hydroxide and aluminum hydroxide; alginic acid; pyrogen-free water; isotonic saline; Ringer's solution; ethyl alcohol and phosphate buffer solutions, as well as other non-toxic compatible substances used in pharmaceutical formulations. Wetting agents, emulsifiers and lubricants such as sodium lauryl sulfate and magnesium stearate, as well as coloring agents, releasing

agents, coating agents, sweetening, flavoring agents, preservatives, stabilizers, extenders, antioxidents, surfactants, solubilizers, lubricants, suspending agents, binders, disintegrating agents, coating materials, etc., can also be present in the composition, according to the judgement of the formulator.

The excipient(s) must be "acceptable" in that the materials are compatible with the other components of the formulation and are not deleterious to the recipient thereof; this includes materials suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problems or complications, commensurate with a reasonable benefit/risk ratio. The compositions of the present invention which include excipients can be formulated according to known methods for the preparation of pharmaceutically useful compositions. Suitable methods are described, for example, in Remington's Pharmaceutical Sciences (19). The proportional ratio of therapeutic agent to excipient will naturally depend on the chemical nature, solubility, and stability of the active ingredient, as well as the dosage contemplated.

The carrier component of the pharmaceutical compositions of the present invention may include polymeric microparticles or nanoparticles of different materials and of very different sizes. Such particles may have a membrane-walled form, in which the core material is concentrated as a reservoir, or a matrix form in which core material is uniformly dispersed. A variety of suitable materials exist ranging from non-degradable polymers, to biodegradable synthetic polymers, to modified natural products such as gums, starches, proteins, fats and waxes (24). The carriers

may also include non-toxic, non-therapeutic components, such as liposomes, starburst polymers, microspheres, microemulsions, nanocapsules or macroemulsions to facilitate formulation, delivery, controlled release or sustained action of the therapeutic composition.

In one embodiment of the present invention, the carrier component of the pharmaceutical composition is a liposome. In an alternate embodiment, the carrier component may be based upon protenoid technology and consist of various amino acids (16).

Liposomes are most frequently prepared from phospholipids, but other molecules of similar molecular shape and dimensions and having both a hydrophobic and a hydrophilic moiety can be used. All such suitable liposome-forming molecules are referred to herein as lipids. One or more naturally occurring and/or synthetic lipid compounds may be used in the preparation of the liposomes.

Liposomes may be anionic, cationic or neutral depending upon the choice of the hydrophilic group. For instance, when a compound with a phosphate or a sulfate group is used, the resulting liposomes will be anionic. When amino-containing lipids are used, the liposomes will have a positive charge, and will be cationic liposomes. In addition, the pharmaceutical compositions of the present invention may include liposome carriers wherein the invasive protein has been incorporated into the liposome bilayer.

Representative suitable phospholipids or lipid compounds for forming liposomes include, but are not limited to, phospholipid-related materials such as phosphatidylcholine (lecithin), lysolecithin, lysophosphatidylethanol-amine, phosphatidylserine, phosphatidylinositol, sphingomyelin, phosphatidyl-ethanolamine (cephalin), cardiolipin, phosphatidic

acid, cerebrosides, dicetylphosphate, phosphatidyl-choline, and dipalmitoyl-phosphatidylglycerol. Additional nonphosphorous-containing lipids include, but are not limited to, stearylamine, dodecylamine, hexadecyl-amine, acetyl palmitate, glycerol ricinoleate, hexadecyl sterate, isopropyl myristate, amphoteric acrylic polymers, fatty acid, fatty acid amides, cholesterol, cholesterol ester, diacylglycerol, diacylglycerolsuccinate, and the like.

In another embodiment of the present invention, the therapeutic agent and the transporting ligand might be incorporated together through a polymeric carrier. For example, the polymeric carrier may be a polymer chain. The list of suitable synthetic polymers includes; poly(ethylene glycol), N-(2-hydroxypropyl)methacrylamide and polyvinyl polymers in particular. Other potential polymeric carriers are polypeptide carriers, such as poly(α amino acids), including poly(α-L-lysine), poly(N^5-hyroxypropyl-L-glutamine), poly(L-aspartic acid). In addition, naturally occurring proteins (albumin, immunoglobulins and lectins), and polysaccharides (dextran and charged derivatives) can be used as carriers. The therapeutic and/or the transporting ligand may be attached to the polymer chain through various reactive side chains that may or may not be degradable *in vivo* (25).

The carrier may be selected or modified to bind the transport enhancer and or the therapeutic agent either through simple absorption, an ionic interaction or covalent linking. Preferably, the carrier is also able to incorporate large amounts of the therapeutic agent in an active form. The carrier component as well as the therapeutic agent associated with the carrier should be stable in the gut environment, but the carrier may also be selected or modified to release the therapeutic agent once it has been

transported across the mucosal barrier. The release of the therapeutic agent may be effectuated by degradative means, such as a cleavable bond, or by degradation of the carrier component. Examples of such release mechanisms may include stabilized Schiff base linkages (26), acid-cleavable linkages (27) or oligonucleotide sequences cleaved by serum factors (28).

The compositions of the present invention are typically formed by attaching the transport enhancer either directly to the therapeutic agent or to a carrier system. Because the bacterial adhesion proteins described in the present invention bind cell receptors, the method of attachment must not prevent the binding of the bacterial protein to the receptor. This can be tested beforehand on *in vitro* systems containing the appropriate receptors, such as membrane preparations or cell systems.

Various conjugation techniques are known in the art, and the following conjugation techniques are provided by way of illustration. Other conjugation techniques can also be used when appropriate as will be appreciated by those skilled in the art. Where the therapeutic agent is a protein, conjugation may be carried out using bifunctional reagents which are capable of reacting with each of the proteins (i.e., the therapeutic protein and the transport enhancer protein) thereby forming a bridge between the two components. Covalent attachment of the transport enhancer to either the therapeutic agent or the carrier system, through either the available amine or carboxy groups of the transport enhancer, may be carried out using suitable conjugation reagents including; glutaraldehyde and cystamine and EDAC. Other known conjugation agents may be used, as long as they provide linkage of the transport factor without

denaturing the protein. One preferred method of conjugation involves thiolation wherein the transport protein is treated with reagents such as N-Succinimidyl 3-(2-pyridyldithio) proprionate(SPDP) to form a disulfide bridge with another sulfhydryl group either in the therapeutic agent or on the carrier. Spacers might also be used and could include polymer chains such as polyethylene glycol, a sugar or a peptide sequence.

Alternatively, the transport enhancer could be attached through a simple absorption method as described in a following Examples. In yet another embodiment, the compositions of the present invention can be in the form of a fusion protein made by recombinant DNA techniques. Thus, one of ordinary skill can duplicate or mimic bacterial proteins which are suitable as transport enhancers. The use of recombinant DNA techniques requires knowledge of the nucleic acid sequence of the polypeptide or protein therapeutic agent to be delivered. The nucleic acid fragment corresponding to the therapeutic agent is linked to a nucleic acid fragment corresponding to the chosen transport enhancer, thereby forming a recombinant molecule. The recombinant molecule is then operably linked to an expression vector and introduced into a host cell to enable expression of a fusion peptide (29) useful as a chimeric molecule in the present invention. When the carrier component of the pharmaceutical composition is also an amino acid sequence, for example a polymer chain, the entire pharmaceutical composition may be produced by recombinant techniques.

The suitability of the resultant pharmaceutical composition as an oral or topical dosage form can be tested following the protocols set forth in the

following Examples. Compositions which are formulated based upon the description of the present invention will be administered to subjects at a dosage range determined by a skilled investigator or attending physician based upon known and accepted parameters. The dosage regimen involved for a particular therapeutic agent may be determined empirically, and making such determinations is within the skill in the art. Prior to administering the agent, it is preferable to determine toxicity levels of the therapeutic agent(s) so as to avoid deleterious effects. Other considerations will include various factors which modify the action of drugs, e.g., the age, condition, body weight, sex and diet of the patient, the nature and severity of the condition as well as any complicating illness, time of administration and other clinical factors. Optimal dosages of the drug of interest can be determined by one of ordinary skill in the art using conventional techniques. As a general rule, the dosage levels will correspond to the accepted and established dosage for the particular therapeutic agent to be delivered, i.e., the dosage will be adjusted to attain clinical equivalence and/or bioequivalence to the parenteral dosage form of the therapeutic agent, or correspond to the dosage that achieves the desired physiological or therapeutic response.

EXAMPLES

EXAMPLE 1

Internalization of INV- and AIL-Transfected Bacteria into the Human Enterocyte

A transfected bacterium which expresses the bacterial adhesion protein on its surface effectively serves as a model for the immobilization of the proteins on the surface of a carrier. For example, the size of a possible microsphere carrier and an *E. coli* bacterium are very similar (approximately 1µm in diameter). Non-transfected *E. coli* serve as a control in the following comparison studies.

To determine if a bacterial coat protein might serve as a transport enhancer, it was first resolved that the protein was able to mediate the adherence, internalization and ultimately transcytosis or transport of transfected bacteria across a layer of polarized human enterocytes. To test this scenario, an *in vitro* model of a cellular layer/barrier was established.

Methods:

Transfection and Maintenance of the Bacteria

Yersinia enterocolitica (8081c), *E. coli* PBR 322 (control plasmid-transfected) and *E. coli* HB101 carrying recombinant plasmids with the *Y. enterocolitica* invasion genes for INV (*E. coli* PVM 101) and AIL (*E. coli* PVM 102), were grown and stored as previously described (7). The construction of the plasmids for the transfection was also performed as described in Miller and Falkow (8).

For the bacteria/cell interaction experiments, *Y. enterocolitica* 8081c was incubated over night in Luria

broth (LB) at room temperature. *E. coli* PBR 322, PVM 101 and PVM 102 were incubated over night in LB, containing 100 µg/ml ampicillin, at 37°C. The approximate bacterial density was then determined by measuring the optical density (O.D.) of the bacterial suspensions and comparing the measurement to a standard curve of O.D. versus bacterial number.

Cell Culture

The Caco-2 cell line (Ciba-Geigy Pharmaceuticals, Horsham, Surrey) was used in the transport studies. The cells were routinely used between passage numbers 95-120, maintained at 37°C under 10% CO_2 in T175 flasks (Falcon Labware, Bedford, MA). Culture medium consisted of Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum, 1% minimum essential medium (MEM) non-essential amino acids, 1000 U/ml penicillin, 100 µg/ml streptomycin and 0.3 mg/ml glutamine. Cell stocks were passaged every five days by briefly washing (x2) with Dulbecco's phosphate buffered saline (PBS) [-Ca^{2+}, -Mg^{2+}], and incubating for ten minutes at 37°C with 0.05% trypsin and 0.53 mM EDTA. Cells were passaged at a ratio of 1:3 and were fed every day except for the first day after passaging. (All solutions were from Gibco, Grand Island, NY).

Non-Polarized Cell Culture

Non-polarized cells were grown on plastic culture dishes. The Caco-2 cells were passaged as described above, diluted into culture medium and then counted on a Neubauer hemocytometer (American Scientific Products, McGaw Park, IL), to determine cell density. One milliliter of the cell suspension containing 1.8 x 10^5 cells was pipetted into each well of a 24-well culture plate (Falcon Labware, Bedford, MA). The

cells were further incubated for ten days prior to the studies.

Determination of Bacterial Adherence and Invasion

All of the culture medium used in the bacterial studies was antibiotic-free. Non-polarized Caco-2 cells, in 24-well Falcon culture plates or Caco-2 monolayers, were washed (x2) in antibiotic-free culture medium 24 hours prior to the experiment. After further incubation over night, the cells were placed in fresh medium and equilibrated for one hour.

The non-polarized cell monolayers were routinely inoculated with approximately 2.5 x 10^5 bacteria per well. The cells were assayed for both surface bound bacteria and invaded/internalized bacteria using known methods (30).

Results:

Bacterial Attachment and Internalization in the Non-Polarized Human Enterocyte

Figure 4 illustrates the effect of invasin on the binding of *E. coli* to the non-polarized human enterocyte Caco-2 cell line and shows that the wild type *Yersinia*, which would be expressing all of the potentially invasive proteins, rapidly adheres to the nonpolarized Caco-2 cell layer. The INV-transfected *E. coli* (closed circles) also demonstrates a rapid surface attachment to the human enterocyte cell line. Levels of surface adhered PVM 101 (INV) are at least 10-fold greater than that of the *Yersinia* bacteria after nine hours of incubation. The *E. coli* control also shows some adherence to the Caco-2 cells, although levels are always 10-fold less than the *Yersinia* or PVM 101. *E. coli* is known to have some

adherent capability in the intestine through the 987P pilus (31).

A major difference occurs in the internalization of the bacteria into the non-polarized cell. Figure 5 illustrates the effect of invasin on the internalization of *E. coli* into the human enterocyte Caco-2 cell line. This internalization is an important prerequisite to transcytosis or delivery across the epithelial barrier. Levels of the internalized *Yersinia* climb rapidly to reach a plateau of 1 x 10^5 CFU/well. Internalized levels of the INV-transfected *E. coli* (closed circles) are much slower to increase but reach 1 x 10^3 CFU/well after nine hours. This is more than 10-fold greater than the internalization of non-transfected *E. coli* which was not greater than 100 CFU/well even after nine hours.

Very similar binding and internalization characteristics are seen for the AIL-transfected *E. coli* bacterium (PVM 102), see Figure 6 and Figure 7. Both the levels of the adhered and the internalized PVM 102, however, are less than the levels mediated by invasin. This could result from the fact that the AIL protein appears to be a later acting protein in the invasion event, as compared to the INV protein. The results demonstrate that both the INV and AIL proteins are able to bind the cells through a receptor expressed on the surface of the human enterocyte which then mediates the uptake of a large bacterial particle (approximately 1 μm) into the cell.

In a separate study the data were reproduced, although only the results at the end of the nine hour incubation are summarized in Figure 8. Again high levels of the *E. coli* control are found adhering to the Caco-2 cells but the levels are less than for any of the bacterium that express the invasive proteins.

The bacteria can be arranged in order of internalization competence as derived from Figure 8: *Yersinia* > PVM 101 (INV) > PVM 102 (AIL) > PBR 302 (non-transfected).

EXAMPLE 2

Receptor-mediated Transcytosis Across the Polarized Human Enterocyte

For an efficient drug delivery system that is dependent on receptor-mediated uptake of pharmaceutical compositions, delivery via transcytosis is important. Receptor-mediated transcytosis can be defined as the trafficking of the ligand and/or the receptor from one membrane domain to the other in an endosome derived from the plasma membrane.

The transfected bacteria were, therefore, tested for their ability to penetrate or pass through the Caco-2 monolayers by transcytosis as described in Example 1.

Methods:

Polarized Cell Culture

Cell culturing was performed substantially in accordance with the methods of Example 1, with the exception that the invasion and binding studies requiring polarized Caco-2 cells were performed on cells grown on a 25 mm diameter Cyclopore® membrane (polyethylene terephtalate), with a pore size of 0.45 µm and a density of 1.6×10^6 pores/cm^2 (from Falcon Labware). The cells were seeded at a cell density of 1.8×10^5 cells/cm^2 insert, with 2.5 ml/domain of culture medium. Cells routinely reached

confluency at five days. The cells were incubated for a total of 21 days prior to use. As the cells grow and divide, they form a confluent monolayer across the insert. Under these conditions, the cells are able to feed from both sides as they do *in vivo*.

To provide for the measurement of bacterial passage across the cell monolayers, the Caco-2 cells were cultured on filter inserts having larger pores. Collagen-coated Transwell-COL filter inserts (nitrocellulose; Costar, Cambridge, MA) were used (average pore size of 3.0 μm and insert diameter of 24 mm). Caco-2 cells were plated at a cell density of 6.6 $x10^4$ cells/cm^2.

Measurement of Monolayer Confluency and Polarity

Prior to any experiment being conducted on the cell monolayers grown on the filter inserts, the monolayers were tested for confluency by measuring for tight junction formation between cells as determined by trans-epithelial electrical resistance (TEER). TEER was determined using an EVOM-F Epithelial Voltohmeter (World Precision Instruments, New Haven, CT) with STX "chopstick" electrodes. The measured resistance was corrected for the area of the filters and was routinely >1000 ohms.cm^2.

The permeability of the monolayers to polyethylene glycol (PEG) (M. wt 4000 Da), inulin (M. wt. 5,200) and dextran (M. wt 70,000 Da) was routinely determined. ^{14}C-labelled PEG 4000 (1 nmol; 2 x 10^5 disintegrations per minute [dpm]), ^{14}C-inulin (1 nmol; 3.1 x 10^4 dpm) (both from Amersham, Arlington Heights, IL) and ^{14}C-dextran (1 nmol; 9 x 10^4 dpm; from New England Nuclear, Boston, MA) were added to the monolayers in culture medium (2.5 ml) for up to 24 hours. Medium (100 μl) from both the apical and basolateral domains was removed after thorough mixing,

aliquoted into XtalScint Ready caps (Beckman, Fullerton, CA) and counted in a Beckman 6000 scintillation counter. The amount of ^{14}C-PEG, ^{14}C-inulin or ^{14}C dextran that had diffused through the monolayer was then calculated. Only the monolayers which demonstrated a TEER > 1000 ohms.cm^2 and a PEG diffusion of <2% in 24 hours were used for both the monolayer characterization and for the bacterial studies.

Formation of intrinsic factor and vitamin B12 complex

The method was derived from Gottlieb et al. (32) adapted by Allen and Mehlman (33). The required amount of ^{57}Co-labelled vitamin B12 from Amersham (CT2; 100-300 µCi/µg) was incubated with a 2-fold molar excess of porcine intrinsic factor (IF) (Sigma, St. Louis, MO). Incubation was in PBS (2 ml) containing 1 mM $CaCl_2$, 0.5 mM $MgCl_2$ (PBS^{++}) with 0.1% bovine serum albumin (BSA) mixing end over end at 4°C for two hours. An equal volume of freshly prepared dextran-coated charcoal, 0.5% charcoal, 0.1% dextran in PBS^{++} at 4°C, was added, vortexed thoroughly and incubated for ten minutes at 4°C. The charcoal was pelleted by centrifugation at 3,000 rpm (1,500 xg) for 15 minutes in an IEC-Centra-8R centrifuge. The supernatant containing the IF-^{57}Co-Vitamin B12 (IF-^{57}Co-VB12) complex was collected for further binding studies. Non-labelled vitamin B12 (VB12) was used in place of ^{57}Co-VB12 to make the IF-VBI2 complex for a determination of non-specific binding.

Binding Studies

Studies were performed to determine the polarity of receptor distribution of the Caco-2 cells on the filter inserts. The studies involved the use of the complexed IF ^{57}Co-VB12 (IF-VB12), ^{125}I-fibronectin (FN;

from ICN, Minneapolis, MN), ^{14}C-taurocholic acid (TA) (54 mCi/mmol) from Amersham and ^{125}I-epidermal growth factor (EGF) (1354 Ci-mmol) also from Amersham. Twenty-four hours prior to the binding studies, the cells were washed (x3) with binding medium (serum-free culture medium with 0.1% BSA, Sigma) and were then further incubated overnight. Immediately prior to the experiment, the medium was replaced again with fresh binding medium, and the cells were incubated for a further hour at 37°C. The cells were then cooled to 4°C for 30 minutes, and the appropriate ligand was added to either the apical or basolateral domains.

^{125}I-FN was added to a final concentration of 86 pM, and for the determination of non-specific binding, a 100-fold molar excess of non-labelled fibronectin was added. IF-^{57}Co-VB12 was present at 100 pM, again with a 100-fold molar excess of non-labelled IF-VB12 for the determination of the non-specific binding. ^{125}I-EGF was present at 80 pM, with and without a 100-fold molar excess of the non-labelled EGF. ^{14}C-TA was present at 400 nM, with and without a 100-fold excess of the non-labelled taurocholic acid.

The incubations were all carried out at 4°C for six hours. To remove the unbound ligand, the cells were washed (x3) at 4°C with PBS. For determination of the γ emitters, the membranes with the cells were cut out of the inserts and counted directly in 12 x 75 mm test tubes in a Cobra 2000 gamma counter (Packard, Meridan, CT). Cells incubated with ^{14}C-TA were solubilized in 0.1 N NaOH (1 ml) and then detected following the addition of 10 ml of Atomlight (New England Nuclear, Cambridge, MA), in a Beckman 6000 scintillation counter.

Determination of Bacterial Invasion of the Monolayers

The TEER of each monolayer was checked immediately prior to the bacterial inoculations with approximately 10^7 bacteria per filter insert, and ^{14}C-PEG 4000 (1 nmol/insert) was also added at this time to monitor monolayer leaking throughout the experiment. Incubation of the cells with the bacteria, was for four hours at 37°C unless otherwise depicted in the figure. The polarized monolayers were routinely evaluated for TEER at each time point, and basolateral medium (100 µl) was removed for the determination of ^{14}C-PEG diffusion.

Adaptations to the protocol of Isberg (30) were used for the determination of bacterial invasion on the polarized cells as follows: at the end of the incubation period on the monolayers, the cells on the Falcon inserts or on the Transwell-COL inserts, were cooled to 4°C before aspirating the medium from both domains. Cells were washed with ice-cold PBS (x5) on either domain, and one milliliter of a 1% Triton X-100 solution in PBS was added and incubated for five minutes at room temperature. Luria broth (1.5 ml) was added to the solubilized cells, which were serially diluted further in LB and plated onto LB agar plates with or without ampicillin for *E. coli* and *Y. enterocolitica*, respectively. Plates were incubated over night, and colonies were counted to determine the total number of bacteria [colony forming Units (CFU)], associated with the cells.

Invasion of the bacteria into the cells of the monolayer was determined by washing the cells with PBS at room temperature, and adding 2.5 ml of medium containing gentamicin sulfate (100 µg/ml) to both domains. After a further 90 minutes at 37°C and washing with PBS (x2), the cells were solubilized and analyzed for CFU as described above.

Determination of Bacterial Passage Across the Monolayers

To study bacterial passage across the monolayer, the incubations were continued for up to 24 hours. To prevent bacterial overgrowth, a "kill" of the apically-located bacteria was performed six hours after bacterial inoculation. Medium in the apical domain was aspirated, and culture medium (2.5 ml), containing gentamicin sulfate (50 µg/ml) was added. After a further incubation for one hour, the apical medium was replaced with culture medium containing gentamicin sulfate (1 µg/ml) and ^{14}C-PEG (1 nmol). The number of bacteria in the basolateral domain of the Transwell-COL inserts was determined at various times. The filter inserts were removed from the wells, transferred to 6-well plates containing pre-equilibrated culture medium (2.5 ml) and further incubated as required. The medium from the used plates was analyzed for both ^{14}C-PEG and total number of bacteria, by determining CFU on agar plates as previously described.

Results:

The Caco-2 cell line is derived from a human colonic tumor and exhibits a morphology consistent with that of the gut epithelium (34). The Caco-2 cells, therefore, provide a generally accepted model for the human enterocyte (35-38). The cells can be grown as a confluent monolayer on plastic cultureware, but under these conditions they are not polarized, i.e., do not have sorted and differentiated domains. Any receptors expressed by the cells, therefore, are distributed over the entire surface of the cell. Alternatively, the cells may be grown as a polarized epithelial-like monolayer on a microporous membrane.

Under these conditions the various receptors are sorted between the two membrane domains, and the cells are a true *in vitro* model of the epithelial lining of the human gut.

The monolayer has tight junctions between the cells which makes the cell monolayer highly impermeable to most molecules having a molecular weight >500 Da (38). The tight junctions separate the apical (lumenal) and basolateral (serosal) domains of the cells (39). In addition, the membrane in each domain is sorted or specific to that domain, such that the receptor population (40) and even the lipids are different in the two domains (41, 42).

The electrical resistance and impermeability of the monolayers is shown in Table 2. After just 12 days in culture, the cells formed confluent monolayers with tight junctions, as demonstrated by the electrical resistance. The electrical resistance does increase somewhat after a further seven days in culture, up to 821 $\Omega.cm^2$.

Table 2
Polarity of the Caco-2 Monolayers

Parameter	Measurement
TEER / 12 days in culture	735.1 ± 17.4 $\Omega.cm^2$
TEER / 19 days in culture	821.6 ± 76.6 $\Omega.cm^2$
	(cm/min)
^{14}C-PEG diffusion Blank	$6.7 \times 10^{-4} \pm 1.16 \times 10^{-5}$
+ Cells	$4.8 \times 10^{-5} \pm 5.76 \times 10^{-6}$
^{14}C-inulin diffusion Blank	$3.04 \times 10^{-4} \pm 8.6 \times 10^{-6}$
+ Cells	$1.97 \times 10^{-6} \pm 1.68 \times 10^{-6}$
^{14}C-dextran diffusion Blank	$5.52 \times 10^{-4} \pm 3.32 \times 10^{-5}$
+Cells	$3.86 \times 10^{-6} \pm 3.0 \times 10^{-6}$

The monolayers were most permeable to the 4000 molecular weight PEG (see Table 2) with a permeability coefficient of 4.8×10^{-5} cm/min. The cells were highly impermeable to a ^{14}C-labelled dextran with a molecular weight of 70,000 Da (a permeability coefficient of 3.86×10^{-6} cm/min). With the Caco-2 cells being impermeable to relatively small molecules, one would expect that they would be impenetrable by relatively large particles such as bacteria.

The polarity of the monolayers used in the studies of invasion proficient bacterial proteins is depicted in Figure 9. The data demonstrate that the

receptor population is sorted according to apical and basolateral membrane domains.

The fibronectin receptor (FN-R) is only found on the basolateral domain. This might be expected of a receptor whose major role is to bind the cell to the extracellular matrix (43). This is of concern, however, since the FN-R is a β_1 integrin receptor, similar to the receptor for the INV protein (4). The epidermal growth factor receptor (EGF-R) is also found predominantly on the basolateral domain (>70%). This is a reasonable outcome because the source of EGF *in vivo* would be from the blood. Similar results with the EGF-R on polarized Caco-2 cells have been demonstrated previously (44).

Two other receptor populations that are normally found on the apical or lumenal side of the gut were also characterized. These were the taurocholic acid receptor (TA-R) (45) and the intrinsic factor receptor (IF-R) (46). IF-R is responsible for the active uptake of vitamin B12 (VB12). Both of these receptors were found predominantly on the apical domain in the *in vitro* model of the polarized human enterocyte. These data agree with previous studies of the polarity of brush border enzymes shown in Caco-2 cells (47).

The data suggest a high degree of polarity of the Caco-2 monolayers on the culture inserts. The cells form an impermeable barrier to most molecules and, therefore, provide a good model for the human gut. Studies to identify invasion proficient bacterial proteins, such as INV and AIL, with this model are reflective of the results one might expect in the human gut.

Bacterial Attachment and Internalization of the Polarized Human Enterocyte

As previously discussed, the polarized *in vitro* model is known to be comparable to the *in vivo* situation as shown by receptor distribution. After bacterial inoculation and as with the non-polarized cells shown for Figures 4-8, relatively high numbers of the non-transfected *E. coli* were seen adhered to the polarized cells, see Figure 10. Again, this may result from some inherent property of the *E. coli*, specifically the 987P pilus (31).

The major effect of the invasive proteins lies in the internalization of the respective bacteria see Figure 11. For the INV- and AIL-transfected *E. coli*, internalized levels of the bacteria were 100-fold and 50-fold greater, respectively, than the non-transfected *E. coli*. In this particular study, the levels of the internalized transfected bacteria were very comparable to those found with the wild-type *Yersinia* bacteria.

The data suggest that the receptors for both the INV and AIL proteins are available on the apical domain of the polarized human enterocyte. This was reassuring following the fibronectin receptor findings (in Figure 9) which suggested that this group of receptors would not be available for binding. After the binding event has occurred through the apical domain, the bacteria are internalized into the cells.

Bacterial Passage Across the Polarized Human Enterocyte

The time course of the trancytosis of the bacteria is shown in Figure 12. The levels of the basolateral-located non-transfected *E. coli* control remained flat throughout the 12 hour study, and were very low. But, both the INV- and AIL-transfected

E. coli are taken up and transcytosed at levels greater than the wild type *Yersinia*, and for the AIL protein the increase is greater than 10-fold. In general, it was found that the AIL protein seemed to mediate the internalization and transcytosis event far more efficiently in polarized human enterocyte Caco-2 cells as compared to non-polarized cells.

The transcytosis mediated by both INV and AIL is quite rapid, but certainly not as quick as the adhesion event. Therefore, any slowness on the part of the proteins to mediate uptake of a particle system will not be detrimental to the system if they also significantly increase the residence time of the protein at the site of uptake through the binding event.

The integrity of the cell monolayer was maintained throughout this study by killing the bacteria in the apical domain were killed after six hours of incubation. Therefore, the bacteria in the basolateral domain represent the bacteria that had been bound and internalized into the enterocytes after the initial six hours of incubation. It should also be noted that the bacteria will continue to divide both inside the cells and after they have crossed the monolayers, and this should be remembered when looking at the total number of bacteria.

To determine the route that the bacteria take across the cell layer, the integrity of the monolayer was checked at the end of every study. ^{14}C-PEG (4000 Da) diffusion was measured as a marker for tight junction integrity between the cells. It was found that the level of PEG diffusion during the 24 hour incubation with the bacteria did not increase over non-inoculated monolayers. This suggests that the bacteria do not cross the monolayers through the tight

junctions nor through a degradation of monolayer integrity. The data suggest that the INV- and AIL-transfected bacteria are able to cross the cells through an internalization and transcytosis event. The finding that the particles crossed the membrane barrier was a novel observation and formed the basis of the current invention.

It has been generally accepted that *Yersinia enterocolitica*, which expresses both INV and AIL, enters the body from the gut through the M cells of the Peyers Patches, (9, 10). This would not be a preferred route for therapeutic delivery. The M cells are the most efficient way to deliver an antigen to the immune system from the gut, and therefore, this route increases the chance of eliciting an immune response to the therapeutic agent. The present data, with the human enterocyte Caco-2 cell line, suggested that a drug delivery system based on INV- or AIL-mediated uptake would also transport a therapeutic agent across the enterocytes, and thereby allow the pharmaceutical composition to reach the systemic circulation. This would increase the potential capacity of the delivery system and decrease or prevent the possible immunologic presentation of the therapeutic agent.

EXAMPLE 3

Expression, purification and testing of the MBP-INV and MBP-AIL fusion proteins

Preparation and purification of bacterial protein

Nucleic acid sequences encoding either the INV or AIL protein, in combination with MBP, were transfected into *E. coli* using known techniques (18). The expressed protein was extracted from the transfected

bacteria by two passes in a French pressure cell at 14,000 p.s.i. The method for the purification of the MBP-INV and MBP-AIL was performed as described by Leong et al. (17) using affinity chromatography with cross-linked amylose (18)).

The amino acid sequence for MBP is illustrated in Figure 3 and SEQ ID NO:3. The amino acid sequence for an exemplary MBP-INV fusion protein is illustrated in Figure 14 and SEQ ID NO:4. The amino acid sequence for an exemplary MBP-AIL fusion protein is illustrated in Figure 15 and SEQ ID NO:5.

In Vitro Assaying of the Fusion Proteins

After purification, the proteins were stored at -80°C, in 10 mM Tris buffer pH 8.0, with 100 mM NaCl and 1 mM EGTA. Assays were established to demonstrate that the proteins were able to bind to the appropriate receptor on the human enterocyte Caco-2 cell after labelling and immobilization of the MBP-INV protein.

Radiolabelling of Bacterial Coat Proteins and MBP-Fusion Protein

Proteins were diluted to a concentration of 500 µg/ml in iodination buffer (100 mM NaH_2PO_4, pH 6.5) and were then microdialyzed over night in iodination buffer. Two Iodobeads (Pierce Chemicals, Rockford, IL) were used per protein and these were prewashed (x2) in iodination buffer, blotted dry and placed in borosilicate tubes. Iodination buffer (100 µl) was added to the beads together with 10 µl of $Na^{125}I$ (carrier free, specific activity 100 mCi/ml, from New England Nuclear). After reacting for five minutes, the protein was added to provide 200 µg/tube. The reaction mixture was mixed, allowed to react for five minutes at room temperature, and was then removed from the Iodobeads. Ten microliters of 1M parahydroxy-

benzoate was added to bind any non-labelling ^{125}I, and the mixture was incubated for a further ten minutes on ice. Separation of the ^{125}I-labelled protein and the unbound ^{125}I was carried out on a PD10 desalting column (Pharmacia, Piscataway, NJ) which had been pre-equilibrated with PBS. Fractions eluted with PBS (500 µl) were collected and assessed for radioactivity in a Cobra 5000 gamma counter (Packard, Downers Grove, IL).

The fractions containing the labelled protein were pooled and then exhaustively dialyzed at 4°C in PBS with 0.02% Tween 20. The dialysate was continually monitored for ^{125}I, until no further non-labelling ^{125}I was removed. The amount of unbound ^{125}I present with the radiolabelled protein was determined by precipitation with a final 6% solution of trichloroacetic acid (TCA). The amount of protein was determined using the BCA protein assay (Pierce Chemicals, Rockford, IL). The final yield of MBP-INV after radiolabelling was 29%. The amount of unbound ^{125}I was 1.5% and the specific activity of the radiolabelled MBP-INV was 3.23×10^6 cpm/µg.

Binding Assay

A conventional binding assay was performed using ^{125}I-labelled MBP-INV, and the specificity of the cell binding with this protein was determined by competing with non-labelled MBP-INV, MBP-AIL and the MBP protein alone. ^{125}I-MBP-INV was added to each well of a 24-well plate containing a confluent monolayer of the Caco-2 cells. The final concentration of the protein was 100 ng/ml (833 pM) and 3.2×10^5 cpm/ml. A 100-fold excess of each competing protein was added as required. The cells were incubated with the proteins for two hours at 37°C under 10% CO_2 in DMEM with 10% fetal bovine serum (FBS). After cooling the cells to 4°C for 30 minutes, the cells were washed (x3) with

PBS containing 0.1% BSA and solubilized in 0.1N NaOH before counting in the Cobra 6000 gamma counter.

Results

The results are summarized in Figure 13. The binding of ^{125}I-labelled MBP-INV was inhibited by more than 70% by the non-labelled MBP-INV, whereas the MBP-AIL protein did not appear to inhibit binding. The control protein MBP, did appear to cause some inhibition of the MBP-INV binding (27%). The results, however, indicate that the INV protein binds the Caco-2 cells through a receptor-specific mechanism. More importantly, the isolated form of the protein retained its binding ability and, therefore, provided a suitable invasion proficient bacterial protein for use in the pharmaceutical compositions of the present invention.

EXAMPLE 4

INV and AIL Proteins with Carrier Component

One embodiment of the pharmaceutical composition of the present invention involves a therapeutic/carrier combination whose uptake is mediated by a transport enhancer, such as the INV or AIL proteins. The MBP-INV protein was associated with fluorescently labelled microspheres and liposomes to evaluate such a delivery system.

Methods:

Coating of Latex Microspheres with Bacterial Proteins

Latex microspheres, labelled with a fluorescent dye (phycoerythrin, PC) and having an average diameter of 0.996 μm, were obtained from Polysciences,

Warrington, PA. The PC-labelled microspheres (2.27 x 10^{10}) were washed (x4) with a 0.1M borate buffer pH 8.5. After each wash, the microspheres were collected by centrifugation at 8,000 rpm for six minutes in an Eppendorf centrifuge.

The latex microspheres were coated with the bacterial coat protein by simple adsorption. The microspheres were resuspended in 300 microliters of 10 mM Tris buffer (pH 8.0) containing 100 mM NaCl, 1 mM EGTA and 400 µg of the MBP-INV protein. A further one milliliter of the borate buffer was then added.

To remove the free or uncoated protein, the microspheres were again centrifuged at 11,000 rpm for ten minutes in the Eppendorf centrifuge, and the supernatant was collected for protein determination in the BCA assay. It was usual that no free protein was found remaining in the supernatant, i.e., all the protein was coating the microspheres. The coated microspheres were subsequently resuspended in the borate buffer (1 ml) with 10 mg/ml BSA, incubated for 30 minutes at room temperature, and then collected by centrifugation. The microspheres were washed (x2) with the borate buffer/BSA (1 ml) before being finally resuspended in PBS (1 ml) containing 10 mg/ml of BSA, 0.1% NaN_3 and 5% glycerol. The microspheres were then stored at 4°C.

Adherence of the INV-Coated Microspheres to Cultured Cells

Two cell lines were used to evaluate the adherence of the bacterial protein/microsphere compositions: the HEp-G2 cell line, (from a human hepatocellular carcinoma cell line from ATCC #HB-8065) and the Caco-2 cell line. The HEp-G2 cell line is epithelial in morphology and is routinely used as an

in vitro cell model of the liver hepatocyte. The cells were plated onto glass coverslips (Baxter, McGaw Park, IL) at a cell density of 1 x 10^5 cells/cm^2 in a 6-well Costar culture plate. The cells were incubated for two days in Dulbecco's minimum essential medium, with 5% FBS and 0.1% non-essential amino acids (all from Gibco), at 37°C and 5% CO_2. DMEM (2 ml) was added to the wells with INV-coated PC-microspheres (2 x 10^8). Control wells were established using uncoated PC-microspheres (2 x 10^8). The cells were further incubated on a rocker at 37°C for two hours before cooling to 4°C and washing (x3) with ice-cold PBS (2 ml). The coverslips were then viewed under a Nikon Optiphot-2 microscope with fluorescence adaptation, and photographs were taken using a Nikon Fx-35WA camera.

Conjugation of MBP-invasin to liposomes

The liposomes were composed of dipalmitoyl-phosphatidylcholine (DPPC):cholesterol (chol):N-glutaryl-dioleoylphosphatidylethanolamine (NG-DOPE) were prepared by sonication. Solvent free lipid films were prepared at a mole ratio of DPPC:chol:NG-DOPE of 2:1:0.1 and contained a trace amount of [^{3}H]-cholesteryl hexadecyl ether (CE) as a marker for total lipid. The lipid films were hydrated in Mes-acetate saline buffer (20 mM Mes, 20 mM NaAcetate, pH 5.5, 0.15 M NaCl) and sonicated to form small unilamellar liposomes. To 0.2 µmol total lipid was added 0.4 mg of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 0.2 mg of N-hydroxysulfosuccinimide (S-NHS), and the samples were mixed for 15 minutes at room temperature. MBP-invasin (0.2 mg) was added and the pH of the suspension adjusted to 8.0 using a small aliquot of 0.4 M $NaHCO_3$ buffer. The sample was then stirred overnight at 4°C.

Unconjugated MBP-invasin was removed from liposomes by centrifuging the samples for 10 minutes at 100,000 x g in an air driven ultracentrifuge. Pelleted liposomes were resuspended in PBS, pH 7.0 and centrifuged twice more to remove unconjugated protein. The conjugated MBP-invasin was determined using the BCA assay, and lipid recovery was quantitated by scintillation counting. The final MBP-invasin:total lipid ratio was between 60 and 100 µg/µmol lipid.

Uptake of liposomes by Caco-2 cells

Dilutions of unconjugated liposomes and MBP-invasin conjugated liposomes were made in RPMI medium (Gibco) and incubated with confluent monolayers of non-polarized Caco-2 cells grown in a 24 well plate for one hour at 37°C. The cells were washed three times with RPMI medium and dissolved by adding 0.1 N NaOH (1 ml) to each well. Dissolved cells (100µl) were used to quantitate cellular protein, while 900 µl of the samples were processed for scintillation counting and lipid quantitation.

Results

A highly visible difference in the adherence of the coated microspheres vs. non-coated microspheres was found on the cells on the coverslips, i.e., the coated microspheres became adherent to the human enterocyte. The effect was observed on both HEp-G2 cells and on the human enterocyte cell line Caco-2. The non-coated microspheres, however, showed no visible adherence to the Caco-2 cells.

The data for the MBP-INV-conjugated liposomes are presented in Figure 16. The results demonstrate an uptake of 5.6-fold greater levels of the MBP-INV-conjugated liposomes over the non-conjugated liposomes

(1.47 nmol/well vs. 0.265 nmol/well). The amount of lipid uptake was found to be concentration dependent.

REFERENCES

1. Clamp, J.R. In: Food Allergy and Intolerance, Edited by J. Brostoff and S.J. Challacombe, pp. 190-205 (1987)
2. Thomson, A.B.R. and Dietschy, J.M. In: Pharmacology of the Intestinal Permeation II, Edited by T.Z. Czaky, 20 (1984)
3. Smith, P.L. et al., *Adv. Drug Del. Rev.* 8, 253-290 (1992)
4. Isberg, R.R. and Leong, J.L., *Cell*, 60, 861-871. (1990)
5. Hart et al., *Journal of Biological Chemistry*, 269, 12468-12474 (1994)
6. Ruoslahti, E., *Annu. Rev. Biochem.*, 57, 375-413 (1988)
7. Isberg, R.R. et al., *Cell*, 50, 769-778 (1987)
8. Miller, V.L. and Falkow, S., *Infection and Immunity*, 56, 1242-1248 (1988)
9. Hanski, C. et al., *Med. Microbiol. Immunol.*, 178, 289-296 (1989)
10. Grutzkau, A. et al., *Gut*, 31, 1011-1015 (1990)
11. Muranashi, S., *Crit. Rev. Ther. Drug Carrier Systems*, 7, 1-33 (1990)
12. Bally et al., U.S. Patent No. 5,171,578
13. Isberg et al., WO 90/12867, published November 1, 1990
14. Jani, P.U. et al., *J. Pharm. Pharmacol.*, 42, 821-826 (1990)
15. Jani, P.U. et al., *Int. J. Pharmaceutics*, 86, 239-246 (1992)
16. Steiner et al., U.S. Patent No. 4,925,673

17. Leong, J.M. et al., *EMBO Journal*, 9, 1979-1989 (1990)
18. Guan, C. et al., *Gene*, 67, 21-30 (1988)
19. *Remington's Pharmaceutical Sciences*, 18TH Edition, A.R. Gennaro, Ed., Mack Publishing Company, (1990)
20. Young, U.B. et al., *Molecular Microbiology*, 4, 1119-1128 (1990)
21. Miller, U.C. et al., *J. Bacteriol.*, 172, 1062-1069 (1990)
22. Beer, K.B. and Miller, V.L., *J. Bacteriol.*, 174, 1360-1369 (1992)
23. Mark et al., U.S. Patent No. 4,518,584.
24. Donbrow, M., In: Microcapsules and Nanoparticles in Medicine and Pharmacy, CRC Press (1992)
25. Hoes, C.J.T. and Feijen, J., In Drug Carrier Systems. Edited by Roerdink, F.H. and Kroon, A.M., 57-109 (1989)
26. Sivam et al., Patent Application No. WO 90/03401
27. Blattler et al., U.S. Patent No. 4,618,492
28. Rutter, U.S. Patent No. 4,769,326
29. *Molecular Cloning, A Laboratory Manual*, Sambrook et al., eds. Cold Spring Harbor Laboratory, 2nd. Ed., Cold Spring Harbor, NY (1989)
30. Isberg, R.R., *Infection and Immunity*, 57, 1998-2008 (1989)
31. Dean, E.A. and Isaacson, R.E., *Infection and Immunity*, 47, 98-105 (1985); *ibid.* 47, 345-348 (1985)
32. Gottlieb et al., *J. Hematology*, 25, 875-886 (1965)
33. Allen and Mehlman, *J. Biol. Chem.*, 248, 3660-3669 (1973)
34. Fogh, J., *J. Nat. Cancer Inst.* 58, 209-214 (1977); *ibid.* 59, 221-226 (1977)
35. Blais, A. et al., *J. Membrane Biol.*, 99, 113-115 (1987)

36. Hidalgo, I. et al., *Gastroenterology*, 96, 736-749 (1989)
37. Wilson, G. et al., *J. Controlled Release*, 11, 25-40 (1990)
38. Rubas, W. et al., *Pharm. Res.*, 10, 113-118 (1993)
39. von Bonsdorff, C.H. et al., *EMBO Journal*, 4, 2781-2792 (1985)
40. Le Bivic, A. et al., *J. Cell Biol.*, 111, 1351-1361 (1990)
41. Simons, K. and van Meer, G., *Biochem.*, 27, 6197-6202 (1988)
42. van't Hof, W. and van Meer, G., *J. Cell Biol.* 111, 977-986 (1990)
43. Ruoslahti, E., *Annual Review of Biochemistry*, 57, 375-413 (1988)
44. Hidalgo, I.J. et al., *Biochem. Biophys. Res. Commun.*, 160, 317-324 (1989)
45. Hidalgo, I.J. and Borchardt, R.T., *Pharm. Res.*, 5, S100 (Abstr) (1989)
46. Dix, C.J. et al., *Biochem. Soc. Trans.* 15, 439-440 (1987)
47. Pinto, M. et al., *Biol. Cell*, 47, 323-330 (1983)
48. Gonnella, P.A., WO 93/20834, published October 28, 1993

The foregoing descriptions of the specific embodiments fully reveal the general nature and applicability of the present invention such that others can readily adapt and/or optimize the teachings and specific embodiments to produce an assortment of pharmaceutical compositions using a variety of therapeutic agents, carrier components and invasive protein transport enhancers. Any such modifications and adaptations are intended to be embraced within the meaning and range of equivalents of the disclosed embodiments. It is also to be understood that the

phraseology and terminology employed herein are for the purpose of description and not of limitation.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Amgen Inc.

(ii) TITLE OF INVENTION: COMPOSITIONS FOR INCREASED BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Amgen Inc.
(B) STREET: 1840 Dehavilland Drive
(C) CITY: Thousand Oaks
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 91320-1789

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3600 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 413..2920

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGTCGTACT GTGGGGAAAA CCGGCGAGAG CGAAGCGGCG GTCCATATAC CCTCCTTAAC 60

TAAGCCAGCG GTTGCTTAGT CGCATTAGAT TAATGCATCG TGAAATGCAG AGAGTCTATT 120

TTATGAGACG AATGTAAACT ATTTTGATAA TAATAATATA TCACAATATA TATATACATG 180

CTAAATATAA CCTGACAATT AAATTAACAA GCTAATATTA CCATGATGAT TTTTTTTTTT 240

TGCATTTCAT TTGTCATTGC TGTTATTTTT AATTTTTTAA TTTTATTTTT GTAAGTTCTG 300

```
CTATTCTATT GTTAGTGTTT GCGAGAGAGA AGAAGTTATT TCTTGTCGCT GTTTTCATTT      360

CTGTTGCTTA AGTAAATATT ACCGCGTTAA TTTATACCTA AGGGGTACAC TA ATG          415
                                                          Met
                                                           1

TAT TCA TTT TTT AAT ACG CTA ACT GTG ACT AAA ATC ATT AGC AGG CTA        463
Tyr Ser Phe Phe Asn Thr Leu Thr Val Thr Lys Ile Ile Ser Arg Leu
               5                  10                  15

ATA TTA TCG ATC GGT TTA ATA TTT GGA ATA TTT ACT TAT GGG TTC TCA        511
Ile Leu Ser Ile Gly Leu Ile Phe Gly Ile Phe Thr Tyr Gly Phe Ser
      20                  25                  30

CAG CAA CAT TAT TTT AAT TCA GAA GCG TTA GAG AAC CCC GCT GAA CAT        559
Gln Gln His Tyr Phe Asn Ser Glu Ala Leu Glu Asn Pro Ala Glu His
     35                  40                  45

AAT GAG GCT TTC AAT AAG ATA ATC AGT ACC GGA ACC AGT CTG GCG GTA        607
Asn Glu Ala Phe Asn Lys Ile Ile Ser Thr Gly Thr Ser Leu Ala Val
 50                  55                  60                  65

TCG GGT AAT GCA TCC AAT ATC ACC AGG TCA ATG GTA AAT GAC GCG GCA        655
Ser Gly Asn Ala Ser Asn Ile Thr Arg Ser Met Val Asn Asp Ala Ala
                 70                  75                  80

AAT CAG GAA GTA AAA CAC TGG TTA AAT AGA TTT GGG ACA ACT CAG GTC        703
Asn Gln Glu Val Lys His Trp Leu Asn Arg Phe Gly Thr Thr Gln Val
             85                  90                  95

AAT GTT AAC TTT GAT AAA AAG TTC TCC CTC AAA GAA AGT TCT CTT GAT        751
Asn Val Asn Phe Asp Lys Lys Phe Ser Leu Lys Glu Ser Ser Leu Asp
        100                 105                 110

TGG CTG TTG CCT TGG TAT GAC TCT GCT TCA TAT GTC TTT TTT AGT CAG        799
Trp Leu Leu Pro Trp Tyr Asp Ser Ala Ser Tyr Val Phe Phe Ser Gln
    115                 120                 125

TTG GGT ATA AGA AAT AAA GAC AGT CGC AAT ACC CTT AAT ATC GGC GCT        847
Leu Gly Ile Arg Asn Lys Asp Ser Arg Asn Thr Leu Asn Ile Gly Ala
130                 135                 140                 145

GGG GTG CGT ACC TTC CAA CAA AGT TGG ATG TAT GGC TTT AAC ACT TCC        895
Gly Val Arg Thr Phe Gln Gln Ser Trp Met Tyr Gly Phe Asn Thr Ser
                150                 155                 160

TAT GAC AAT GAT ATG ACT GGG CAC AAT CAT CGT ATT GGC GTG GGT GCA        943
Tyr Asp Asn Asp Met Thr Gly His Asn His Arg Ile Gly Val Gly Ala
            165                 170                 175

GAA GCC TGG ACT GAT TAT TTA CAA TTA TCG GCC AAT GGT TAT TTT CGC        991
Glu Ala Trp Thr Asp Tyr Leu Gln Leu Ser Ala Asn Gly Tyr Phe Arg
        180                 185                 190

CTC AAT GGT TGG CAT CAA TCT CGT GAT TTC GCG GAC TAT AAT GAG CGC       1039
Leu Asn Gly Trp His Gln Ser Arg Asp Pne Ala Asp Tyr Asn Glu Arg
    195                 200                 205
```

```
CCG GCA AGC GGG GGC GAC ATT CAC GTC AAA GCG TAT TTA CCT GCG CTG      1087
Pro Ala Ser Gly Gly Asp Ile His Val Lys Ala Tyr Leu Pro Ala Leu
210                 215                 220                 225

CCA CAA TTG GGC GGG AAA TTA AAA TAT GAG CAG TAC CGT GGT GAG CGG      1135
Pro Gln Leu Gly Gly Lys Leu Lys Tyr Glu Gln Tyr Arg Gly Glu Arg
                230                 235                 240

GTG GCT TTA TTT GGT AAA GAT AAC CTG CAA AGT AAC CCT TAT GCG GTG      1183
Val Ala Leu Phe Gly Lys Asp Asn Leu Gln Ser Asn Pro Tyr Ala Val
            245                 250                 255

ACC ACA GGG CTT ATT TAT ACG CCG ATC CCC TTC ATT ACA CTG GGG GTC      1231
Thr Thr Gly Leu Ile Tyr Thr Pro Ile Pro Phe Ile Thr Leu Gly Val
        260                 265                 270

GAT CAA CGA ATG GGA AAA AGT CGG CAG CAT GAA ATA CAA TGG AAC TTA      1279
Asp Gln Arg Met Gly Lys Ser Arg Gln His Glu Ile Gln Trp Asn Leu
    275                 280                 285

CAA ATG GAT TAT CGC CTC GGC GAA AGT TTT CGT TCG CAG TTT AGC CCC      1327
Gln Met Asp Tyr Arg Leu Gly Glu Ser Phe Arg Ser Gln Phe Ser Pro
290                 295                 300                 305

GCA GTG GTG GCC GGA ACT CGT TTA CTG GCT GAG AGC CGT TAT AAT CTG      1375
Ala Val Val Ala Gly Thr Arg Leu Leu Ala Glu Ser Arg Tyr Asn Leu
                310                 315                 320

GTT GAG CGC AAT CCA AAT ATT GTT CTG GAA TAC CAA AAA CAG AAT ACT      1423
Val Glu Arg Asn Pro Asn Ile Val Leu Glu Tyr Gln Lys Gln Asn Thr
            325                 330                 335

ATC AAA TTG GCA TTT TCA CCC GCC GTA CTC TCC GGC CTG CCG GGG CAG      1471
Ile Lys Leu Ala Phe Ser Pro Ala Val Leu Ser Gly Leu Pro Gly Gln
        340                 345                 350

GTT TAT TCC GTT AGT GCA CAA ATA CAG TCT CAA TCT GCA CTA CAA CGT      1519
Val Tyr Ser Val Ser Ala Gln Ile Gln Ser Gln Ser Ala Leu Gln Arg
    355                 360                 365

ATT CTC TGG AAT GAT GCG CAA TGG GTT GCT GCC GGC GGC AAA TTA ATA      1567
Ile Leu Trp Asn Asp Ala Gln Trp Val Ala Ala Gly Gly Lys Leu Ile
370                 375                 380                 385

CCC GTC AGT GCA ACA GAT TAC AAT GTG GTC TTA CCG CCT TAT AAA CCG      1615
Pro Val Ser Ala Thr Asp Tyr Asn Val Val Leu Pro Pro Tyr Lys Pro
                390                 395                 400

ATG GCA CCA GCG AGT CGT ACT GTG GGG AAA ACC GGC GAG AGC GAA GCG      1663
Met Ala Pro Ala Ser Arg Thr Val Gly Lys Thr Gly Glu Ser Glu Ala
            405                 410                 415

GCG GTC AAT ACC TAT ACC CTC AGC GCC ACG GCT ATC GAT AAC CAC GGC      1711
Ala Val Asn Thr Tyr Thr Leu Ser Ala Thr Ala Ile Asp Asn His Gly
        420                 425                 430

AAT AGT TCT AAT CCA GCT ACG TTG ACC GTT ATT GTG CAG CAA CCT CAG      1759
Asn Ser Ser Asn Pro Ala Thr Leu Thr Val Ile Val Gln Gln Pro Gln
    435                 440                 445
```

```
TTC GTT ATT ACC TCG GAA GTG ACT GAT GAT GGT GCG CTT GCT GAT GGC      1807
Phe Val Ile Thr Ser Glu Val Thr Asp Asp Gly Ala Leu Ala Asp Gly
450                 455                 460                 465

AGG ACT CCC ATC ACG GTG AAA TTT ACA GTG ACT AAT ATT GAT AGT ACG      1855
Arg Thr Pro Ile Thr Val Lys Phe Thr Val Thr Asn Ile Asp Ser Thr
                470                 475                 480

CCG GTT GCC GAG CAA GAG GGG GTG ATA ACC ACC AGT AAT GGT GCG CTT      1903
Pro Val Ala Glu Gln Glu Gly Val Ile Thr Thr Ser Asn Gly Ala Leu
            485                 490                 495

CCC AGT AAA GTC ACA AAA AAA ACC GAT GCA CAG GGT GTG ATA AGC ATT      1951
Pro Ser Lys Val Thr Lys Lys Thr Asp Ala Gln Gly Val Ile Ser Ile
        500                 505                 510

GCA TTA ACT AGC TTC ACT GTT GGG GTG TCA GTC GTC ACT TTA GAT ATT      1999
Ala Leu Thr Ser Phe Thr Val Gly Val Ser Val Val Thr Leu Asp Ile
    515                 520                 525

CAG GGG CAA CAG GCT ACT GTT GAT GTA CGA TTT GCC GTG CTG CCG CCA      2047
Gln Gly Gln Gln Ala Thr Val Asp Val Arg Phe Ala Val Leu Pro Pro
530                 535                 540                 545

GAT GTC ACA AAC TCA AGT TTT AAC GTT TCT CCA TCT GAT ATT GTT GCC      2095
Asp Val Thr Asn Ser Ser Phe Asn Val Ser Pro Ser Asp Ile Val Ala
                550                 555                 560

GAT GGC TCC ATG CAG TCG ATA CTC ACC TTT GTT CCG CGT AAT AAA AAT      2143
Asp Gly Ser Met Gln Ser Ile Leu Thr Phe Val Pro Arg Asn Lys Asn
            565                 570                 575

AAT GAG TTT GTC AGT GGG ATA ACA GAT CTT GAA TTT ATA CAA AGT GGT      2191
Asn Glu Phe Val Ser Gly Ile Thr Asp Leu Glu Phe Ile Gln Ser Gly
        580                 585                 590

GTT CCG GTA ACT ATT AGT CCG GTA ACC GAA AAT GCT GAC AAC TAT ACC      2239
Val Pro Val Thr Ile Ser Pro Val Thr Glu Asn Ala Asp Asn Tyr Thr
    595                 600                 605

GCC AGT GTG GTG GGA AAT TCG GTA GGA GAT GTC GAT ATT ACG CCG CAG      2287
Ala Ser Val Val Gly Asn Ser Val Gly Asp Val Asp Ile Thr Pro Gln
610                 615                 620                 625

GTG GGG GGG GAA TCA CTA GAC TTG TTG CAG AAA AGA ATC ACC CTG TAC      2335
Val Gly Gly Glu Ser Leu Asp Leu Leu Gln Lys Arg Ile Thr Leu Tyr
                630                 635                 640

CCA GTA CCG AAG ATA ACC GGC ATT AAC GTG AAT GGT GAG CAA TTT GCC      2383
Pro Val Pro Lys Ile Thr Gly Ile Asn Val Asn Gly Glu Gln Phe Ala
            645                 650                 655

ACA GAT AAA GGC TTC CCG AAA ACT ACC TTT AAT AAA GCC ACG TTC CAA      2431
Thr Asp Lys Gly Phe Pro Lys Thr Thr Phe Asn Lys Ala Thr Phe Gln
        660                 665                 670
```

```
TTG GTG ATG AAT GAC GAT GTG GCG AAT AAT ACT CAA TAT GAC TGG ACA      2479
Leu Val Met Asn Asp Asp Val Ala Asn Asn Thr Gln Tyr Asp Trp Thr
    675                 680                 685

TCA TCC TAT GCG GCC AGT GCG CCG GTT GAT AAT CAG GGT AAA GTC AAT      2527
Ser Ser Tyr Ala Ala Ser Ala Pro Val Asp Asn Gln Gly Lys Val Asn
690                 695                 700                 705

ATT GCC TAT AAA ACC TAT GGT AGC ACC GTC ACT GTG ACG GCA AAA AGT      2575
Ile Ala Tyr Lys Thr Tyr Gly Ser Thr Val Thr Val Thr Ala Lys Ser
                710                 715                 720

AAA AAA TTC CCG AGT TAT ACG GCA ACA TAT CAA TTC AAA CCT AAT TTG      2623
Lys Lys Phe Pro Ser Tyr Thr Ala Thr Tyr Gln Phe Lys Pro Asn Leu
            725                 730                 735

TGG GTG TTC TCC GGC ACC ATG TCA CTG CAA TCA AGT GTC GAG GCG AGT      2671
Trp Val Phe Ser Gly Thr Met Ser Leu Gln Ser Ser Val Glu Ala Ser
        740                 745                 750

CGA AAT TGC CAG CGC ACT GAT TTT ACT GCG CTG ATC GAG TCC GCA CGC      2719
Arg Asn Cys Gln Arg Thr Asp Phe Thr Ala Leu Ile Glu Ser Ala Arg
    755                 760                 765

GCC AGT AAT GGT TCG CGT TCA CCA GAC GGT ACT CTG TGG GGA GAG TGG      2767
Ala Ser Asn Gly Ser Arg Ser Pro Asp Gly Thr Leu Trp Gly Glu Trp
770                 775                 780                 785

GGA AGT TTG GCA ACC TAT GAT AGC GCT GAG TGG CCA TCG GGT AAC TAT      2815
Gly Ser Leu Ala Thr Tyr Asp Ser Ala Glu Trp Pro Ser Gly Asn Tyr
                790                 795                 800

TGG ACT AAA AAG ACC AGT ACA GAT TTT GTC ACT ATG GAT ATG ACC ACC      2863
Trp Thr Lys Lys Thr Ser Thr Asp Phe Val Thr Met Asp Met Thr Thr
            805                 810                 815

GGT GAC ATA CCA ACA TCT GCG GCT ACG GCG TAT CCG CTG TGT GCG GAG      2911
Gly Asp Ile Pro Thr Ser Ala Ala Thr Ala Tyr Pro Leu Cys Ala Glu
        820                 825                 830

CCG CAA TAGTGCTAAA TACCAATCTT GCGGCCCAGC AAACTGGCAC CTTTAGCGTG      2967
Pro Gln
    835

ACCATCTGGC CCATACAGTG ATTGGCCGTG GCGCGTATTC AAAACCGCCA GCGCCTGAGT   3027

GTTATGCTCA ATATGCTGTT GCAGCAAAAG CCCGTTATGC AGGTTGCCGT AGCGCACCGT   3087

TCGGCCAGTT CCAAAATACG CTGCCAGCGC TCAGCTAGCG CAGGAACGTT GCTGTAGGGC   3147

GCTTGAATAT TTATGTTTTT TTCGGTGGTG AGCCGGGTCT GGTCCAGATA AGCCAAGGTG   3207

CCAAAATTGA ACTTTTTTGT TCAGTGACGC CTTGCAACAC GATACCTTGA ATCCGACCGG   3267

AGCACAGCAG TTGCTGCTCT TGTGCTACTA CGGTTTTCAG GGATTCAAGC AGTTCCAGTT   3327

GCTGGTCCAG ATTAGTTTGT AATCTTTCCA CCACCACCTA TCCTTTTACG GTTAATAATT   3387

TTACGGTCAA CGATTGTTGT GACGTTTAGC TATTCTTCAG GTCATCGGCA ACATTTTTGA   3447
```

GCAAGGCATC GGCAATTTTA CCGGCGTCCA TGGTCAGTTG GCCTGAACGG ATCGCCTGTT 3507

TTAAGGTTTC GACACGTTCT ACATTGATGT CCTGGCTGCC CGGTTGCATC AATTTTGCCT 3567

GCGCGTCGCT CAATTTAACC TCAGTACCAC TTA 3600

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2220 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 536..1024

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGATCACATC ATCAATACCG AAGCCCAAGA GATTAGCCAG TGTGCCAGAA AAATGGCTCG 60

ATGGGACGTT GGTGGAAGGA AGCAAATATT GCTTACGGCA CGAAAACGCA TGATAGATGA 120

GCTTCAGATG TATTTGCCAG GACTGGGAAG TCACGTGGGT AATTACTGTG ACATCCAGTA 180

ATAAAACAGA GCCTCTATTA AAGGAGCTTC CCAATTTGAA ATCAGAAAAA TTACATCATA 240

AACATGGGTG TCCAGAAGTC AGTCGGCGAT ATATCCATTT AAAGAGCATT GAGCTATGAC 300

CAGTATTCAT CAACTACAGA ACAAAAATAC AGGAATAAGT GACTGATGGG ATAAAGCTGA 360

GGTAAGCTCA CAGTACTGTA TCAATATCCA TATTTACATA TATATCATGG ATTTGGCATT 420

ATATCATCAG CCATGTCAGT GATATGGTTA TTGTATTAGT ATTGTTATAA CAATCTGGAT 480

TATTTTTATG AAAAAGACAT TACTAGCTAG TTCTCTAATA GCCTGTTTAT CAATT GCG 538
Ala
1

TCT GTT AAT GTG TAC GCT GCG AGT GAA AGT AGT ATT TCT ATT GGT TAT 586
Ser Val Asn Val Tyr Ala Ala Ser Glu Ser Ser Ile Ser Ile Gly Tyr
5 10 15

GCG CAA AGC CAT GTA AAA GAA AAT GGG TAT ACA TTG GAT AAT GAC CCT 634
Ala Gln Ser His Val Lys Glu Asn Gly Tyr Thr Leu Asp Asn Asp Pro
20 25 30

AAA GGT TTT AAC CTG AAG TAC CGT TAT GAA CTC GAT GAT AAC TGG GGA 682
Lys Gly Phe Asn Leu Lys Tyr Arg Tyr Glu Leu Asp Asp Asn Trp Gly
35 40 45

```
GTA ATA GGT TCG TTT GCT TAT ACT CAT CAG GGA TAT GAT TTC TTC TAT        730
Val Ile Gly Ser Phe Ala Tyr Thr His Gln Gly Tyr Asp Phe Phe Tyr
 50                  55                  60                  65

GGC AGT AAT AAG TTT GGT CAT GGT GAT GTT GAT TAC TAT TCA GTA ACA        778
Gly Ser Asn Lys Phe Gly His Gly Asp Val Asp Tyr Tyr Ser Val Thr
                 70                  75                  80

ATG GGG CCA TCT TTC CGC ATC AAC GAA TAT GTT AGC CTT TAT GGA TTA        826
Met Gly Pro Ser Phe Arg Ile Asn Glu Tyr Val Ser Leu Tyr Gly Leu
             85                  90                  95

CTG GGG GCC GCT CAT GGA AAG GTT AAG GCA TCT GTA TTT GAT GAA TCA        874
Leu Gly Ala Ala His Gly Lys Val Lys Ala Ser Val Phe Asp Glu Ser
         100                 105                 110

ATC AGT GCA AGT AAG ACG TCA ATG GCA TAC GGG GCA GGG GTG CAA TTC        922
Ile Ser Ala Ser Lys Thr Ser Met Ala Tyr Gly Ala Gly Val Gln Phe
     115                 120                 125

AAC CCA CTT CCA AAT TTT GTC ATT GAC GCT TCA TAT GAA TAC TCC AAA        970
Asn Pro Leu Pro Asn Phe Val Ile Asp Ala Ser Tyr Glu Tyr Ser Lys
130                 135                 140                 145

CTC GAT AGC ATA AAA GTT GGC ACC TGG ATG CTT GGT GCA GGG TAT CGA       1018
Leu Asp Ser Ile Lys Val Gly Thr Trp Met Leu Gly Ala Gly Tyr Arg
                150                 155                 160

TTC TAATCATCTC AGATAGTGAA AACCCACCTG AGTGAAGTGA ACCCCATTTA            1071
Phe

TTGGACACTT TTCCTGGCGG TTGACATGGC CTGATTTCGG TACTGCACCG GACTCAGGCC     1131

GTTTAATTTT ACTTTGATCC TTTCGTTGTT GTAGTAATGG ATATACTCAT CCACCGCTTT     1191

TTTCAGTTGT TCTACATCTT CGTATTTTTC ATTGTGCCAG CATTCAGTCT TCAGCAGACC     1251

AAAAAAGTTT TCTATCACAG CATTATCCAG GCAGTTGCCC TTGCGCGACA TACTTTGCTT     1311

TACTTCGCCA GACCCCAGCC TTTTCTTATA GCTTGCCATC TGATATTGCC AGCCCTGATC     1371

CGAGTGAAGT ACAGGTTCAT CGCCTGAGTT CAACTTCTGT AGCGCATCAT CAAGCATTTT     1431

ATCAATCAGG TTCATTCCGG GATGCGTATC CATCTGCCAG GCAACGACTT CGCTGTTATA     1491

CAGATCCAGC ACGGGTGACA GATACAGCTT TTTACCCCTG ACGTTGAACT CGGTCACATC     1551

GTTACCCACT TCTGGTTAGG GGCTTCGGCA GTAAATTTTC GAGCAAGTAT ATTAGGGACC     1611

ACTTTACCGT AGGCACCCTG ATATGACTGA TATTTTTTAC GACGCAAGTT AGATGCAAGC     1671

TGCTGTTGCC GCATGAGTTT TCGTACGGTT TTATGGTTAA GACTCCCGCC CTCATTGCGT     1731

AGGGCCAGCG TTATTCTGCG GTAACCATAG CGACCTTTAT GATGGTGAAA CAGGGTTTTT     1791

ATTCTTTGTT TCTCATCCGC ATAACTCTCT TCACGACCAC TGGATTTTAC CTGCCAGTAG     1851

AAGGTGCTGC GCGGAAGACC GGCCACGTAA AGCAAGGTCG CCAGTTTATA CAGATGCCTT     1911
```

AATTCAGTGA TTATTCGCGT TTTTTCCGCT GCTTCTCTTA CAGGTGGTAT TCACTGAGTG 1971

CCACCGATAA TGCGCAGGCA AAGTCATTAA CGACCCCCGC CGCTCACCCT GAGCATGGTC 2031

GTTGATGGCT TTTATATTTT CCATAGAGCA GAGGATGATT CTTTATGTCC CGAGTGAACT 2091

GGGGTGAACG GTTATCCCGG TTTGCCGCTG AATGGCAACG GACGGGAATA TCCCCTAAAG 2151

AGTGGTGTGA GAGAGAAGGT TATTCGTGGG GAACAGCGAA AGCGTATATT TCGATAAAAG 2211

CAGCGAAAG 2220

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6545 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3630..4820

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCTTTTCCG TGGTATGACC AGAACATAAA GTTTTTGCTG CCCCAACGCC GGTGTCAGGC 60

GCATAACGCC TTCCAGCCGA TCTGCACTCA TCACGCCTGG TTCCTAGTAG GTGAAATAAC 120

TGCTGGGGAA GAAGGCTGAT GGGGTCATGT TCTGATCAAG AATCAGCACG TTCGGCGCAA 180

AAACGCTGGT TTGTTTGTTC ACTTCGCTGG TCAGCGTCAG GGTCAGTTCG CCAATGTTTG 240

CCGGGACGCT GTACGCAGCA ACCGGACCAC TGATGCCGGG AACGTTCAGT TGTTGGCCGC 300

CGGTCGCCAG TTGGGTGGTC TGGGTTTTAG ATTGATCGAC CGGTGTCCAG GTGAGTTGTT 360

GCAGCGCAGC AGATGGAATG GCTGGCGCGT CGCTGGTGTT TTGCGGTACG TAGTTAACAT 420

CGGCAAGGCT AATTCCAGGC GCGCTTGCCA GTAACCCTGC TGATAAACAG AGGACGATGA 480

GACTTTTATT CATTTTCATT GTTTTCACCT CAAAATCTGG AGCTCAGCGG TAGCCAGGCA 540

ATAGCGCGCT AAACCCGATA ATCAGAGGGG CTTTCGCCCC TTCAGATAAT GACAACCTGT 600

TTTTATGCCG GATGCGGCGT AAACGCCTTA TCCGGCCTAC ATTTGACAGC CGTTGTAGGC 660

CTGATAAGAC GCGCAAGCGT CGCATCAGGC GTTGGTTGCC GAATGCGGCG TAAACGCCTT 720

ATCCGGCCCA GGTTTTGCTA TTACCACCAG ATTTCCATCT GGGCACCGAA GGTCCACTCG 780

TCGCTGTCGC CACGACCGAA GCTGCCGCCG TTGAAATCAG CAGGAACGGC TTTGCCGAAG 840

TTCGCGTTGT TATCAGCGTT ACCGGTGTAG TCGTAACCCC ATTTCTCATC CCACTTGGCG 900

TAGGTTGCGA AGACACGAAT AGCCGGGCGT GACCAGATGC TGTCGCCAGC CTGCCATTGT 960

TGTGCGAGGG TAATTTTGTA CTGATTGTTC TTGTCGCCGG TGCGCTGGGA TTCGACGTTG 1020

TCGTAGCCGA TTTCCATCAC GGTGCTCATG ATTGGCGTCC ACTTGTACAT CGGGCGAATA 1080

CCGACGGTCC ACCACTTGGT GCCGTTGTCG TTATCCCAGT TGATATCCTG GTACATACCC 1140

ACGTACATCA TGTCCCAGTT GTCGCCCATG GAGATCGCAC CGTGGTCGAG GATACGCAGC 1200

ATGTGACCGT TGTTGTTGAT ATTGTAGGCA AATTTTTCGT TATCAAATGC AACGCCAGAA 1260

CCCTGCGACA GCCCTTTACC CTGCGAGGTC ATCGAGTCAG TAGCGTACTG AACAACAAAC 1320

TTGTTAAAGC CCTTCAGGAC ACTCTGAGTA TGTTCAGCAG TGAATAACCA GCCGTCTTTC 1380

GATGCGCCAT CAACCAGACG ATAGTTATCA CGCAAGTTGG CACGACCGTA GTCGACACCC 1440

AGTTCTAATG TGCCGCCCGG GTTGATTTCC ATCTGCGCTA AACGCACATC GAAAACGTCG 1500

TTCGCGGTTT CGTTGGTATA GTCATAAATA TTGTTGCTGG CGAAAGAGGA AGAACCACCA 1560

GCTTCAGAGG AGCGGGTTGC TGCCAGAGAG AGTTTACCGA AGCCAACATC GATGTTTTCC 1620

AGACCGGCAC CAGGACCAGA AATATCCCAG TAGTAGAAGT CGATCATATG AACGTCATGA 1680

CGTTGGTAGA AGCGCTTACC TGCCCAGATG GTGGAGCCTG GCAGCCATTC GATCAGGTTT 1740

TTACCCTGCA CGTTTGCTTC ACGGAAGGCC GGATCGGTAG CTTCCCAGTC ATTCTGTTGT 1800

GCGACGGAAT AGGCCACGTT AGTGTCGAAA TAGAAGCTCT TATCGCCCTC TTTCCACACT 1860

TCCTGACCCA ATTTTAATTC AGCATAAGTT TCACATTCGT TGCCAAGACG GTATTTACTT 1920

TGAGCACCGG TAGTCTGGAA ACACTGTTGT TCACCGCCGC TACCTGTCCA ACCAATACCG 1980

GAACGTGCAT AGCCGTGGAA ATCAACAGCC ATTGCCTGAG CAGACATTAC GCCCGCTGCG 2040

ACGGCAACCG CCAGAGGAAG TTTGCGCAGA GTAATCATCA TTCTATCTCC TGAGTCATTG 2100

CTTTTCTTTT TTCACATCAC CTGTGACAGG CTTTGTGTGT TTTGTGGGGT GCTTAAACGC 2160

CCGGCTCCTT ATGCAGTCGA CGACATGCAG TGCCATCCTC ACGGAACAGA TGGCAACGCT 2220

CTGGCGGCAG GCCGATAGCG AATGTGGCAC CTTCTTCTAC CAACACCACG TCGTTCTGGC 2280

GGTACACCAG GTTTTGACGA ATGGAAGGGA TCTGGATATG GATTTGAGTT TCGTTGCCGA 2340

GTTGCTCGAC GACCTGAACT TCACCCTCAA GGATGACGTC AGCGATATCA CTCGGCAGTA 2400

GATGTTCCGG GCGAATACCC AGCGACATAT TGGCTCCAAC CTGGACATCA CGGCTTTCAA 2460

CTGGCAGCCA GACTTGCTGA CGATTTGGCA TCGGCAGCTC CACCTGCACT TGATCGATTG 2520

CGGTGGCGGT CACTTTTACC GGCAGGAGTT CATCTTTGGC GAACCGATAA ATCCGGCGAC 2580

AAAACGGTCT GCCGGATAGT GGTACAGCTA GCGGTTTCCC AACCTGCGCC ACGCGACCGG 2640

```
CGTCCAGCAC CACGATTTTG TCGGCCAGCG TCATCGCTTC GACCTGATCG TGGGTGACGT    2700

AAATCATTGT GCGGCCCAGG CGTTTATGCA GACGGGAGAT TTCGATACGC ATTTGCACAC    2760

GCAGTGCAGC ATCGAGGTTG GAGAGCGGTT CATCGAGCAA AAATACGCTT GGCTCGGCCA    2820

CCAGCGTACG GCCAATCGCC ACACGCTGAC GCTGACCACC GGAGAGCGCT TTCGGTTTGC    2880

GATCCAGCAA ATGCGCCAGT TGTAGCACTT CCGCCACCTG GTTAACGCGT TGGTTAATCA    2940

CCTCTTTTTT TGCGCCAGCA GGTTTCAGGC CAAATGACAT GTTTTCTGCT ACTGACAGGT    3000

GGGGATAGAG CGCGTAAGAC TGAAACACCA TACCAACGCC GCGTTCTGCT GGCGGAGTGT    3060

CATTCATCCG TTTCTCACCG ATGAACAGGT CGCCGCTGGT GATCGTCTCA AGCCCGGCAA    3120

TCATGCGCAG TAAAGTCGAT TTACCGCAGC CAGACGGTCC GACAAACACC ACGAATTCAC    3180

CTTCATGGAT ATCGAGATTG ATATCTTTCG ATACCACGAC CTCGCCCCAG GCTTTCGTTA    3240

CATTTTGCAG CTGTACGCTC GCCATGCCCT TCTCCCTTTG TAACAACCTG TCATCGACAG    3300

CAACATTCAT GATGGGCTGA CTATGCGTCA TCAGGAGATG GCTTAAATCC TCCACCCCCT    3360

GGCTTTTTTA TGGGGGAGGA GGCGGGAGGA TGAGAACACG GCTTCTGTGA ACTAAACCGA    3420

GGTCATGTAA GGAATTTCGT GATGTTGCTT GCAAAAATCG TGGCGATTTT ATGTGCGCAT    3480

CTCCACATTA CCGCCAATTC TGTAACAGAG ATCACACAAA GCGACGGTGG GGCGTAGGGG    3540

CAAGGAGGAT GGAAAGAGGT TGCCGTATAA AGAAACTAGA GTCCGTTTAG GTGTTTTCAC    3600

GAGCACTTCA CCAACAAGGA CCATAGATT ATG AAA ATA AAA ACA GGT GCA CGC       3653
                               Met Lys Ile Lys Thr Gly Ala Arg
                                 1               5

ATC CTC GCA TTA TCC GCA TTA ACG ACG ATG ATG TTT TCC GCC TCG GCT       3701
Ile Leu Ala Leu Ser Ala Leu Thr Thr Met Met Phe Ser Ala Ser Ala
     10                  15                  20

CTC GCC AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG ATT AAC GGC GAT       3749
Leu Ala Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp
 25                  30                  35                  40

AAA GGC TAT AAC GGT CTC GCT GAA GTC GGT AAG AAA TTC GAG AAA GAT       3797
Lys Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp
                45                  50                  55

ACC GGA ATT AAA GTC ACC GTT GAG CAT CCG GAT AAA CTG GAA GAG AAA       3845
Thr Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys
            60                  65                  70

TTC CCA CAG GTT GCG GCA ACT GGC GAT GGC CCT GAC ATT ATC TTC TGG       3893
Phe Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp
        75                  80                  85
```

```
GCA CAC GAC CGC TTT GGT GGC TAC GCT CAA TCT GGC CTG TTG GCT GAA      3941
Ala His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu
 90                  95                 100

ATC ACC CCG GAC AAA GCG TTC CAG GAC AAG CTG TAT CCG TTT ACC TGG      3989
Ile Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp
105                 110                 115                 120

GAT GCC GTA CGT TAC AAC GGC AAG CTG ATT GCT TAC CCG ATC GCT GTT      4037
Asp Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val
                125                 130                 135

GAA GCG TTA TCG CTG ATT TAT AAC AAA GAT CTG CTG CCG AAC CCG CCA      4085
Glu Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro
            140                 145                 150

AAA ACC TGG GAA GAG ATC CCG GCG CTG GAT AAA GAA CTG AAA GCG AAA      4133
Lys Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys
        155                 160                 165

GGT AAG AGC GCG CTG ATG TTC AAC CTG CAA GAA CCG TAC TTC ACC TGG      4181
Gly Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp
    170                 175                 180

CCG CTG ATT GCT GCT GAC GGG GGT TAT GCG TTC AAG TAT GAA AAC GGC      4229
Pro Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly
185                 190                 195                 200

AAG TAC GAC ATT AAA GAC GTG GGC GTG GAT AAC GCT GGC GCG AAA GCG      4277
Lys Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala
                205                 210                 215

GGT CTG ACC TTC CTG GTT GAC CTG ATT AAA AAC AAA CAC ATG AAT GCA      4325
Gly Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala
            220                 225                 230

GAC ACC GAT TAC TCC ATC GCA GAA GCT GCC TTT AAT AAA GGC GAA ACA      4373
Asp Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr
        235                 240                 245

GCG ATG ACC ATC AAC GGC CCG TGG GCA TGG TCC AAC ATC GAC ACC AGC      4421
Ala Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser
    250                 255                 260

AAA GTG AAT TAT GGT GTA ACG GTA CTG CCG ACC TTC AAG GGT CAA CCA      4469
Lys Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro
265                 270                 275                 280

TCC AAA CCG TTC GTT GGC GTG CTG AGC GCA GGT ATT AAC GCC GCC AGT      4517
Ser Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser
                285                 290                 295

CCG AAC AAA GAG CTG GCG AAA GAG TTC CTC GAA AAC TAT CTG CTG ACT      4565
Pro Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr
            300                 305                 310

GAT GAA GGT CTG GAA GCG GTT AAT AAA GAC AAA CCG CTG GGT GCC GTA      4613
Asp Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val
        315                 320                 325
```

```
GCG CTG AAG TCT TAC GAG GAA GAG TTG GCG AAA GAT CCA CGT ATT GCC      4661
Ala Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala
    330                 335                 340

GCC ACC ATG GAA AAC GCC CAG AAA GGT GAA ATC ATG CCG AAC ATC CCG      4709
Ala Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro
345                 350                 355                 360

CAG ATG TCC GCT TTC TGG TAT GCC GTG CGT ACT GCG GTG ATC AAC GCC      4757
Gln Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala
                365                 370                 375

GCC AGC GGT CGT CAG ACT GTC GAT GAA GCC CTG AAA GAC GCG CAG ACT      4805
Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr
            380                 385                 390

CGT ATC ACC AAG TAATGCTGTG AAATGCCGGA TGCGGCGTGA ACGCCTTGTC          4857
Arg Ile Thr Lys
        395

CGGCCTACAA AACCGAAACG TATGTAGGCC TGATAAGACG CGTCAGCGTC GCATCAGGCA    4917

GTTGTTGTCG GATAAGGCGT GAAAGCCTTA TCCGTCCTGG AATGAGGAAG AACCCCATGG    4977

ATGTCATTAA AAAGAAACAT TGGTGGCAAA GCGACGCGCT GAAATGGTCA GTGCTAGGTC    5037

TGCTCGGCCT GCTGGTGGGT TACCTTGTTG TTTTAATGTA CGCACAAGGG GAATACCTGT    5097

TCGCCATTAC CACGCTGATA TTGAGTTCAG CGGGGCTGTA TATTTTCGCC AATCGTAAAG    5157

CCTACGCCTG GCGCTATGTT TACCCGGGAA TGGCTGGAAT GGGATTATTC GTCCTCTTCC    5217

CTCTGGTCTG CACCATCGCC ATTGCCTTCA CCAACTACAG CAGCACTAAC CAGCTGACTT    5277

TTGAACGTGC GCAGGAAGTG TTGTTAGATC GCTCCTGGCA AGCAGGCAAA ACCTATAACT    5337

TTGGTCTTTA CCCGGCGGGC GATGAGTGGC AACTGGCGCT CAGCGACGGC GAAACCGGCA    5397

AAAATTACCT CTCCGACGCT TTTAAATTTG GCGGCGAGCA AAAACTGCAA CTGAAAGAAA    5457

CGACCGCCCA GCCCGAAGGC GAACGCGCGA ATCTGCGCGT GATTACCCAG AATCGTCAGG    5517

CGCTGAGTGA CATTACCGCC ATTCTGCCGG ATGGCAACAA AGTGATGATG AGCTCCCTGC    5577

GCCAGTTTTC TGGCACGCAG CCGCTCTACA CACTCGACGG TGACGGCACG TTGACGAATA    5637

ATCAGAGCGG CGTGAAATAT CGTCCGAATA ACCAAATTGG CTTTTACCAG TCCATTACCG    5697

CCGACGGCAA CTGGGGTGAT GAAAAGCTAA GCCCCGGTTA CACCGTGACC ACCGGCTGGA    5757

AAAACTTTAC CCGCGTCTTT ACCGACGAAG GCATTCAGAA ACCGTTCCTC GCCATTTTCG    5817

TCTGGACCGT GGTGTTCTCG CTGATCACTG TCTTTTTAAC GGTGGCGGTC GGCATGGTTC    5877

TGGCGTGTCT GGTGCAGTGG GAAGCGTTGC GCGGCAAAGC GGTCTATCGC GTCCTGCTGA    5937

TTCTGCCCTA CGCGGTGCCA TCGTTCATTT CAATCTTGAT TTTCAAAGGG TTGTTTAACC    5997
```

AGAGCTTCGG TGAAATCAAC ATGATGTTGA GCGCGCTGTT TGGCGTGAAG CCCGCCTGGT 6057

TCAGCGATCC GACCACCGCC CGCACGATGC TAATTATCGT CAATACCTGG CTGGGTTATC 6117

CGTACATGAT GATCCTCTGC ATGGGCTTGC TGAAAGCGAT TCCGGACGAT TTGTATGAAG 6177

CCTCAGCAAT GGATGGCGCA GGTCCGTTCC AGAACTTCTT TAAGATTACG CTGCCGCTGC 6237

TGATTAAACC GCTGACGCCG CTGATGATCG CCAGCTTCGC CTTTAACTTT AACAACTTCG 6297

TGCTGATTCA ACTGTTAACC AACGGCGGCC CGGATCGTCT TGGCACGACC ACGCCAGCCG 6357

GTTATACCGA CCTGCTTGTT AACTACACCT ACCGCATCGC TTTTGAAGGC GGCGGGGGTC 6417

AGGACTTCGG TCTGGCGGCA GCAATTGCCA CGCTGATCTT CCTGCTGGTG GGTGCGCTGG 6477

CGATAGTGAA CCTGAAAGCC ACGCGAATGA AGTTTGATTA AGGGAGATAA CAAAAATGGC 6537

AATGGTCC 6545

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 588 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45

Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65 70 75 80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125

```
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130                 135                 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145                 150                 155                 160

Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
                165                 170                 175

Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
            180                 185                 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
        195                 200                 205

Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
    210                 215                 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225                 230                 235                 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
                245                 250                 255

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
            260                 265                 270

Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
        275                 280                 285

Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
    290                 295                 300

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305                 310                 315                 320

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
                325                 330                 335

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
            340                 345                 350

Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
        355                 360                 365

Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
    370                 375                 380

Glu Ala Leu Lys Asp Ala Gln Thr Arg Ile Thr Lys Val Pro Thr Leu
385                 390                 395                 400

Thr Gly Ile Leu Val Asn Gly Gln Asn Phe Ala Thr Asp Lys Gly Phe
                405                 410                 415

Pro Lys Thr Ile Phe Lys Asn Ala Thr Phe Gln Leu Gln Met Asp Asn
            420                 425                 430

Asp Val Ala Asn Asn Thr Gln Tyr Glu Trp Ser Ser Ser Phe Thr Pro
        435                 440                 445
```

Asn Val Ser Val Asn Asp Gln Gly Gln Val Thr Ile Thr Tyr Gln Thr
450 455 460

Tyr Ser Glu Val Ala Val Thr Ala Lys Ser Lys Lys Phe Pro Ser Tyr
465 470 475 480

Ser Val Ser Tyr Arg Phe Tyr Pro Asn Arg Trp Ile Tyr Asp Gly Gly
485 490 495

Arg Ser Leu Val Ser Ser Leu Glu Ala Ser Arg Gln Cys Gln Gly Ser
500 505 510

Asp Met Ser Ala Val Leu Glu Ser Ser Arg Ala Thr Asn Gly Thr Arg
515 520 525

Ala Pro Asp Gly Thr Leu Trp Gly Glu Trp Gly Ser Leu Thr Ala Tyr
530 535 540

Ser Ser Asp Trp Gln Ser Gly Glu Tyr Trp Val Lys Lys Thr Ser Thr
545 550 555 560

Asp Phe Glu Thr Met Asn Met Asp Thr Gly Ala Leu Gln Pro Gly Pro
565 570 575

Ala Tyr Leu Ala Phe Pro Leu Cys Ala Leu Ser Ile
580 585

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 568 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45

Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65 70 75 80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125

Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160

Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
165 170 175

Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205

Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270

Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285

Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
340 345 350

Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
355 360 365

Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
370 375 380

Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser Val Pro Gly Arg
385 390 395 400

Gly Ser Ile Glu Gly Arg Ala Ser Val Asn Val Tyr Ala Ala Ser Glu
405 410 415

Ser Ser Ile Ser Ile Gly Tyr Ala Gln Ser His Val Lys Glu Asn Gly
420 425 430

Tyr Thr Leu Asp Asn Asp Pro Lys Gly Phe Asn Leu Lys Tyr Arg Tyr
435 440 445

Glu Leu Asp Asp Asn Trp Gly Val Ile Gly Ser Phe Ala Tyr Thr His
450 455 460

Gln Gly Tyr Asp Phe Phe Tyr Gly Ser Asn Lys Phe Gly His Gly Asp
465 470 475 480

Val Asp Tyr Tyr Ser Val Thr Met Gly Pro Ser Phe Arg Ile Asn Glu
485 490 495

Tyr Val Ser Leu Tyr Gly Leu Leu Gly Ala Ala His Gly Lys Val Lys
500 505 510

Ala Ser Val Phe Asp Glu Ser Ile Ser Ala Ser Lys Thr Ser Met Ala
515 520 525

Tyr Gly Ala Gly Val Gln Phe Asn Pro Leu Pro Asn Phe Val Ile Asp
530 535 540

Ala Ser Tyr Glu Tyr Ser Lys Leu Asp Ser Ile Lys Val Gly Thr Trp
545 550 555 560

Met Leu Gly Ala Gly Tyr Arg Phe
565

CLAIMS

What is claimed is:

1. A therapeutic delivery system for the delivery of a therapeutic agent, comprising:
a) a therapeutic agent; and
b) an invasion proficient bacterial protein which transports the composition across the gastrointestinal membrane barrier via transcytosis, and thereby increases the systemic bioavailability of said therapeutic agent.

2. The delivery system according to Claim 1, wherein transcytosis via said bacterial protein increases the systemic bioavailability of said therapeutic agent by 5-fold to 100-fold.

3. The delivery system according to Claim 1, wherein said bacterial protein is invasin protein.

4. The delivery system according to Claim 1, wherein said bacterial protein is attachment-invasion-locus protein.

5. The delivery system according to Claim 1, further comprising a carrier component.

6. The delivery system according to Claim 5, wherein said carrier component is selected from the group consisting of liposomes, and polymer-based particles.

7. The delivery system according to Claim 1, wherein said therapeutic agent and said invasion proficient bacterial protein are linked by a degradable peptide sequence.

8. A pharmaceutical composition comprising:

a) a therapeutic agent;

b) an invasion proficient bacterial protein which transports the composition across the gastrointestinal tract; and

c) a carrier component.

9. The composition according to Claim 8, wherein said bacterial protein is invasin or attachment-invasion-locus protein or a fragment thereof.

10. The composition according to Claim 8, wherein said therapeutic agent and said bacterial protein are linked by a degradable peptide sequence.

11. The composition according to Claim 8, wherein said carrier is selected from the group consisting of a liposome and a polymer particle.

12. A pharmaceutical composition comprising: a fusion protein including a therapeutic moiety and an invasion proficient bacterial protein to effect delivery of the composition across the gastrointestinal tract.

13. The composition according to Claim 12, wherein said bacterial protein is invasin protein.

14. The composition according to Claim 12, wherein said bacterial protein is attachment-invasion-locus protein.

15. The composition according to Claim 12, further comprising a carrier component.

16. The composition according to Claim 15, wherein said carrier component is selected from the group consisting of liposomes and polymer-based particles.

17. The composition according to Claim 12, wherein said therapeutic moiety and said invasion proficient bacterial protein are linked by a degradable peptide sequence.

18. A method of delivering a therapeutic agent through the gastrointestinal membrane barrier, comprising: orally administering a pharmaceutical composition comprising a therapeutic agent and an invasion proficient bacterial protein.

19. The method according to Claim 18, wherein said invasion protein is invasin protein.

20. The method according to Claim 18, wherein said invasion protein is attachment-invasion-locus protein.

21. The method according to Claim 18, wherein said pharmaceutical composition further comprises a carrier component.

22. The method according to Claim 18, wherein said pharmaceutical composition comprises a fusion protein including said therapeutic agent and said invasion protein.

23. A pharmaceutical composition comprising: a fusion protein comprising a therapeutic agent, an invasion proficient bacterial protein to effect delivery of the composition across the gastrointestinal tract and a carrier component.

Figur 1

```
    GAGTCGTACT GTGGGGAAAA CCGGCGAGAG CGAAGCGGCG GTCCATATAC
-1  ----------+----------+----------+----------+----------+   50

    CCTCCTTAAC TAAGCCAGCG GTTGCTTAGT CGCATTAGAT TAATGCATCG
51  ----------+----------+----------+----------+----------+  100

    TGAAATGCAG AGAGTCTATT TTATGAGACG AATGTAAACT ATTTTGATAA
101 ----------+----------+----------+----------+----------+  150

    TAATAATATA TCACAATATA TATATACATG CTAAATATAA CCTGACAATT
151 ----------+----------+----------+----------+----------+  200

    AAATTAACAA GCTAATATTA CCATGATGAT TTTTTTTTTT TGCATTTCAT
201 ----------+----------+----------+----------+----------+  250

    TTGTCATTGC TGTTATTTTT AATTTTTTAA TTTTATTTTT GTAAGTTCTG
251 ----------+----------+----------+----------+----------+  300

    CTATTCTATT GTTAGTGTTT GCGAGAGAGA AGAAGTTATT TCTTGTCGCT
301 ----------+----------+----------+----------+----------+  350

    GTTTTCATTT CTGTTGCTTA AGTAAATATT ACCGCGTTAA TTTATACCTA
351 ----------+----------+----------+----------+----------+  400

    AGGGGTACAC TAATGTATTC ATTTTTTAAT ACGCTAACTG TGACTAAAAT
401 ----------+----------+----------+----------+----------+  450
                M  Y  S   F  F  N   T  L  T  V   T  K  I

    CATTAGCAGG CTAATATTAT CGATCGGTTT AATATTTGGA ATATTTACTT
451 ----------+----------+----------+----------+----------+  500
     I  S  R   L  I  L  S   I  G  L   I  F  G   I  F  T  Y

    ATGGGTTCTC ACAGCAACAT TATTTTAATT CAGAAGCGTT AGAGAACCCC
501 ----------+----------+----------+----------+----------+  550
      G  F  S   Q  Q  H   Y  F  N  S   E  A  L   E  N  P

    GCTGAACATA ATGAGGCTTT CAATAAGATA ATCAGTACCG GAACCAGTCT
551 ----------+----------+----------+----------+----------+  600
     A  E  H  N   E  A  F   N  K  I   I  S  T  G   T  S  L

    GGCGGTATCG GGTAATGCAT CCAATATCAC CAGGTCAATG GTAAATGACG
601 ----------+----------+----------+----------+----------+  650
      A  V  S   G  N  A  S   N  I  T   R  S  M   V  N  D  A

    CGGCAAATCA GGAAGTAAAA CACTGGTTAA ATAGATTTGG GACAACTCAG
651 ----------+----------+----------+----------+----------+  700
      A  N  Q   E  V  K   H  W  L  N   R  F  G   T  T  Q

    GTCAATGTTA ACTTTGATAA AAAGTTCTCC CTCAAAGAAA GTTCTCTTGA
701 ----------+----------+----------+----------+----------  750
     V  N  V  N   F  D  K   K  F  S   L  K  E  S   S  L  D

    TTGGCTGTTG CCTTGGTATG ACTCTGCTTC ATATGTCTTT TTTAGTCAGT
751 ----------+----------+----------+----------+----------+  800
      W  L  L   P  W  Y  D   S  A  S   Y  V  F   F  S  Q  L
```

Figure 1 (continued)

```
      TGGGTATAAG AAATAAAGAC AGTCGCAATA CCCTTAATAT CGGCGCTGGG
801   ----------+----------+----------+----------+----------+  850
        G  I  R   N  K  D   S  R  N  T   L  N  I   G  A  G

      GTGCGTACCT TCCAACAAAG TTGGATGTAT GGCTTTAACA CTTCCTATGA
851   ----------+----------+----------+----------+----------+  900
      V  R  T  F   Q  Q  S   W  M  Y   G  F  N  T   S  Y  D

      CAATGATATG ACTGGGCACA ATCATCGTAT TGGCGTGGGT GCAGAAGCCT
901   ----------+----------+----------+----------+----------+  950
       N  D  M   T  G  H  N   H  R  I   G  V  G   A  E  A  W

      GGACTGATTA TTTACAATTA TCGGCCAATG GTTATTTTCG CCTCAATGGT
951   ----------+----------+----------+----------+----------+ 1000
        T  D  Y   L  Q  L   S  A  N  G   Y  F  R   L  N  G

      TGGCATCAAT CTCGTGATTT CGCGGACTAT AATGAGCGCC CGGCAAGCGG
1001  ----------+----------+----------+----------+----------+ 1050
      W  H  Q  S   R  D  F   A  D  Y   N  E  R  P   A  S  G

      GGGCGACATT CACGTCAAAG CGTATTTACC TGCGCTGCCA CAATTGGGCG
1051  ----------+----------+----------+----------+----------+ 1100
       G  D  I   H  V  K  A   Y  L  P   A  L  P   Q  L  G  G

      GGAAATTAAA ATATGAGCAG TACCGTGGTG AGCGGGTGGC TTTATTTGGT
1101  ----------+----------+----------+----------+----------+ 1150
        K  L  K   Y  E  Q   Y  R  G  E   R  V  A   L  F  G

      AAAGATAACC TGCAAAGTAA CCCTTATGCG GTGACCACAG GGCTTATTTA
1151  ----------+----------+----------+----------+----------+ 1200
      K  D  N  L   Q  S  N   P  Y  A   V  T  T  G   L  I  Y

      TACGCCGATC CCCTTCATTA CACTGGGGGT CGATCAACGA ATGGGAAAAA
1201  ----------+----------+----------+----------+----------+ 1250
       T  P  I   P  F  I  T   L  G  V   D  Q  R   M  G  K  S

      GTCGGCAGCA TGAAATACAA TGGAACTTAC AAATGGATTA TCGCCTCGGC
1251  ----------+----------+----------+----------+----------+ 1300
        R  Q  H   E  I  Q   W  N  L  Q   M  D  Y   R  L  G

      GAAAGTTTTC GTTCGCAGTT TAGCCCCGCA GTGGTGGCCG GAACTCGTTT
1301  ----------+----------+----------+----------+----------+ 1350
      E  S  F  R   S  Q  F   S  P  A   V  V  A  G   T  R  L

      ACTGGCTGAG AGCCGTTATA ATCTGGTTGA GCGCAATCCA AATATTGTTC
1351  ----------+----------+----------+----------+----------+ 1400
       L  A  E   S  R  Y  N   L  V  E   R  N  P   N  I  V  L

      TGGAATACCA AAAACAGAAT ACTATCAAAT TGGCATTTTC ACCCGCCGTA
1401  ----------+----------+----------+----------+----------+ 1450
        E  Y  Q   K  Q  N   T  I  K  L   A  F  S   P  A  V

      CTCTCCGGCC TGCCGGGGCA GGTTTATTCC GTTAGTGCAC AAATACAGTC
1451  ----------+----------+----------+----------+----------+ 1500
      L  S  G  L   P  G  Q   V  Y  S   V  S  A  Q   I  Q  S
```

Figure 1 (continued)

```
      TCAATCTGCA CTACAACGTA TTCTCTGGAA TGATGCGCAA TGGGTTGCTG
1501  ----------+----------+----------+----------+----------+  1550
        Q  S  A   L  Q  R  I   L  W  N   D  A  Q   W  V  A  A

      CCGGCGGCAA ATTAATACCC GTCAGTGCAA CAGATTACAA TGTGGTCTTA
1551  ----------+----------+----------+----------+----------+  1600
        G  G  K   L  I  P   V  S  A  T   D  Y  N   V  V  L

      CCGCCTTATA AACCGATGGC ACCAGCGAGT CGTACTGTGG GGAAAACCGG
1601  ----------+----------+----------+----------+----------+  1650
        P  P  Y   K  P  M  A   P  A  S   R  T  V  G   K  T  G

      CGAGAGCGAA GCGGCGGTCA ATACCTATAC CCTCAGCGCC ACGGCTATCG
1651  ----------+----------+----------+----------+----------+  1700
        E  S  E   A  A  V  N   T  Y  T   L  S  A   T  A  I  D

      ATAACCACGG CAATAGTTCT AATCCAGCTA CGTTGACCGT TATTGTGCAG
1701  ----------+----------+----------+----------+----------+  1750
        N  H  G   N  S  S   N  P  A  T   L  T  V   I  V  Q

      CAACCTCAGT TCGTTATTAC CTCGGAAGTG ACTGATGATG GTGCGCTTGC
1751  ----------+----------+----------+----------+----------+  1800
        Q  P  Q   F  V  I  T   S  E  V   T  D  D  G   A  L  A

      TGATGGCAGG ACTCCCATCA CGGTGAAATT TACAGTGACT AATATTGATA
1801  ----------+----------+----------+----------+----------+  1850
        D  G  R   T  P  I  T   V  K  F   T  V  T   N  I  D  S

      GTACGCCGGT TGCCGAGCAA GAGGGGGTGA TAACCACCAG TAATGGTGCG
1851  ----------+----------+----------+----------+----------+  1900
        T  P  V   A  E  Q   E  G  V  I   T  T  S   N  G  A

      CTTCCCAGTA AAGTCACAAA AAAAACCGAT GCACAGGGTG TGATAAGCAT
1901  ----------+----------+----------+----------+----------+  1950
        L  P  S   K  V  T  K   K  T  D   A  Q  G  V   I  S  I

      TGCATTAACT AGCTTCACTG TTGGGGTGTC AGTCGTCACT TTAGATATTC
1951  ----------+----------+----------+----------+----------+  2000
        A  L  T   S  F  T  V   G  V  S   V  V  T   L  D  I  Q

      AGGGGCAACA GGCTACTGTT GATGTACGAT TTGCCGTGCT GCCGCCAGAT
2001  ----------+----------+----------+----------+----------+  2050
        G  Q  Q   A  T  V   D  V  R  F   A  V  L   P  P  D

      GTCACAAACT CAAGTTTTAA CGTTTCTCCA TCTGATATTG TTGCCGATGG
2051  ----------+----------+----------+----------+----------+  2100
        V  T  N   S   S  F  N   V  S  P   S  D  I  V   A  D  G

      CTCCATGCAG TCGATACTCA CCTTTGTTCC GCGTAATAAA AATAATGAGT
2101  ----------+----------+----------+----------+----------+  2150
        S  M  Q   S  I  L  T   F  V  P   R  N  K   N  N  E  F

      TTGTCAGTGG GATAACAGAT CTTGAATTTA TACAAAGTGG TGTTCCGGTA
2151  ----------+----------+----------+----------+----------+  2200
        V  S  G   I  T  D   L  E  F  I   Q  S  G   V  P  V
```

Figure 1 (continued)

```
      ACTATTAGTC CGGTAACCGA AAATGCTGAC AACTATACCG CCAGTGTGGT
2201  ---------+ ---------+ ---------+ ---------+ ---------+  2250
      T  I  S  P   V  T  E   N  A  D   N  Y  T  A   S  V  V

      GGGAAATTCG GTAGGAGATG TCGATATTAC GCCGCAGGTG GGGGGGGAAT
2251  ---------+ ---------+ ---------+ ---------+ ---------+  2300
      G  N  S   V  G  D  V   D  I  T   P  Q  V   G  G  E  S

      CACTAGACTT GTTGCAGAAA AGAATCACCC TGTACCCAGT ACCGAAGATA
2301  ---------+ ---------+ ---------+ ---------+ ---------+  2350
        L  D  L   L  Q  K   R  I  T  L   Y  P  V   P  K  I

      ACCGGCATTA ACGTGAATGG TGAGCAATTT GCCACAGATA AAGGCTTCCC
2351  ---------+ ---------+ ---------+ ---------+ ---------+  2400
      T  G  I  N   V  N  G   E  Q  F   A  T  D  K   G  F  P

      GAAAACTACC TTTAATAAAG CCACGTTCCA ATTGGTGATG AATGACGATG
2401  ---------+ ---------+ ---------+ ---------+ ---------+  2450
      K  T  T   F  N  K  A   T  F  Q   L  V  M   N  D  D  V

      TGGCGAATAA TACTCAATAT GACTGGACAT CATCCTATGC GGCCAGTGCG
2451  ---------+ ---------+ ---------+ ---------+ ---------+  2500
        A  N  N   T  Q  Y   D  W  T  S   S  Y  A   A  S  A

      CCGGTTGATA ATCAGGGTAA AGTCAATATT GCCTATAAAA CCTATGGTAG
2501  ---------+ ---------+ ---------+ ---------+ ---------+  2550
      P  V  D  N   Q  G  K   V  N  I   A  Y  K  T   Y  G  S

      CACCGTCACT GTGACGGCAA AAAGTAAAAA ATTCCCGAGT TATACGGCAA
2551  ---------+ ---------+ ---------+ ---------+ ---------+  2600
      T  V  T   V  T  A  K   S  K  K   F  P  S   Y  T  A  T

      CATATCAATT CAAACCTAAT TTGTGGGTGT TCTCCGGCAC CATGTCACTG
2601  ---------+ ---------+ ---------+ ---------+ ---------+  2650
        Y  Q  F   K  P  N   L  W  V  F   S  G  T   M  S  L

      CAATCAAGTG TCGAGGCGAG TCGAAATTGC CAGCGCACTG ATTTTACTGC
2651  ---------+ ---------+ ---------+ ---------+ ---------+  2700
      Q  S  S  V   E  A  S   R  N  C   Q  R  T  D   F  T  A

      GCTGATCGAG TCCGCACGCG CCAGTAATGG TTCGCGTTCA CCAGACGGTA
2701  ---------+ ---------+ ---------+ ---------+ ---------+  2750
        L  I  E   S  A  R  A   S  N  G   S  R  S   P  D  G  T

      CTCTGTGGGG AGAGTGGGGA AGTTTGGCAA CCTATGATAG CGCTGAGTGG
2751  ---------+ ---------+ ---------+ ---------+ ---------+  2800
        L  W  G   E  W  G   S  L  A  T   Y  D  S   A  E  W

      CCATCGGGTA ACTATTGGAC TAAAAAGACC AGTACAGATT TTGTCACTAT
2801  ---------+ ---------+ ---------+ ---------+ ---------+  2850
      P  S  G  N   Y  W  T   K  K  T   S  T  D  F   V  T  M

      GGATATGACC ACCGGTGACA TACCAACATC TGCGGCTACG GCGTATCCGC
2851  ---------+ ---------+ ---------+ ---------+ ---------+  2900
      D  M  T   T  G  D  I   P  T  S   A  A  T   A  Y  P  L
```

Figure 1 (continued)

```
                TGTGTGCGGA GCCGCAATAG TGCTAAATAC CAATCTTGCG GCCCAGCAAA
2901            ----------+ ----------+ ----------+ ----------+ ----------+ 2950
                  C   A   E   P   Q   *

                CTGGCACCTT TAGCGTGACC ATCTGGCCCA TACAGTGATT GGCCGTGGCG
2951            ----------+ ----------+ ----------+ ----------+ ----------+ 3000

                CGTATTCAAA ACCGCCAGCG CCTGAGTGTT ATGCTCAATA TGCTGTTGCA
3001            ----------+ ----------+ ----------+ ----------+ ----------+ 3050

                GCAAAAGCCC GTTATGCAGG TTGCCGTAGC GCACCGTTCG GCCAGTTCCA
3051            ----------+ ----------+ ----------+ ----------+ ----------+ 3100

                AAATACGCTG CCAGCGCTCA GCTAGCGCAG GAACGTTGCT GTAGGGCGCT
3101            ----------+ ----------+ ----------+ ----------+ ----------+ 3150

                TGAATATTTA TGTTTTTTTC GGTGGTGAGC CGGGTCTGGT CCAGATAAGC
3151            ----------+ ----------+ ----------+ ----------+ ----------+ 3200

                CAAGGTGCCA AAATTGAACT TTTTTGTTCA GTGACGCCTT GCAACACGAT
3201            ----------+ ----------+ ----------+ ----------+ ----------+ 3250

                ACCTTGAATC CGACCGGAGC ACAGCAGTTG CTGCTCTTGT GCTACTACGG
3251            ----------+ ----------+ ----------+ ----------+ ----------+ 3300

                TTTTCAGGGA TTCAAGCAGT TCCAGTTGCT GGTCCAGATT AGTTTGTAAT
3301            ----------+ ----------+ ----------+ ----------+ ----------+ 3350

                CTTTCCACCA CCACCTATCC TTTTACGGTT AATAATTTTA CGGTCAACGA
3351            ----------+ ----------+ ----------+ ----------+ ----------+ 3400

                TTGTTGTGAC GTTTAGCTAT TCTTCAGGTC ATCGGCAACA TTTTTGAGCA
3401            ----------+ ----------+ ----------+ ----------+ ----------+ 3450

                AGGCATCGGC AATTTTACCG GCGTCCATGG TCAGTTGGCC TGAACGGATC
3451            ----------+ ----------+ ----------+ ----------+ ----------+ 3500

                GCCTGTTTTA AGGTTTCGAC ACGTTCTACA TTGATGTCCT GGCTGCCCGG
3501            ----------+ ----------+ ----------+ ----------+ ----------+ 3550

                TTGCATCAAT TTTGCCTGCG CGTCGCTCAA TTTAACCTCA GTACCACTTA
3551            ----------+ ----------+ ----------+ ----------+ ----------+ 3600
```

Figure 2

```
    GGATCACATC ATCAATACCG AAGCCCAAGA GATTAGCCAG TGTGCCAGAA
1   ---------+ ---------+ ---------+ ---------+ ---------+  50

    AAATGGCTCG ATGGGACGTT GGTGGAAGGA AGCAAATATT GCTTACGGCA
51  ---------+ ---------+ ---------+ ---------+ ---------+ 100

    CGAAAACGCA TGATAGATGA GCTTCAGATG TATTTGCCAG GACTGGGAAG
101 ---------+ ---------+ ---------+ ---------+ ---------+ 150

    TCACGTGGGT AATTACTGTG ACATCCAGTA ATAAAACAGA GCCTCTATTA
151 ---------+ ---------+ ---------+ ---------+ ---------+ 200

    AAGGAGCTTC CCAATTTGAA ATCAGAAAAA TTACATCATA AACATGGGTG
201 ---------+ ---------+ ---------+ ---------+ ---------+ 250

    TCCAGAAGTC AGTCGGCGAT ATATCCATTT AAAGAGCATT GAGCTATGAC
251 ---------+ ---------+ ---------+ ---------+ ---------+ 300

    CAGTATTCAT CAACTACAGA ACAAAAATAC AGGAATAAGT GACTGATGGG
301 ---------+ ---------+ ---------+ ---------+ ---------+ 350

    ATAAAGCTGA GGTAAGCTCA CAGTACTGTA TCAATATCCA TATTTACATA
351 ---------+ ---------+ ---------+ ---------+ ---------+ 400

    TATATCATGG ATTTGGCATT ATATCATCAG CCATGTCAGT GATATGGTTA
401 ---------+ ---------+ ---------+ ---------+ ---------+ 450

    TTGTATTAGT ATTGTTATAA CAATCTGGAT TATTTTTATG AAAAAGACAT
451 ---------+ ---------+ ---------+ ---------+ ---------+ 500

    TACTAGCTAG TTCTCTAATA GCCTGTTTAT CAATTGCGTC TGTTAATGTG
501 ---------+ ---------+ ---------+ ---------+ ---------+ 550
                                        A  S   V  N  V

    TACGCTGCGA GTGAAAGTAG TATTTCTATT GGTTATGCGC AAAGCCATGT
551 ---------+ ---------+ ---------+ ---------+ ---------+ 600
    Y  A  A  S   E  S  S   I  S  I   G  Y  A  Q   S  H  V

    AAAAGAAAAT GGGTATACAT TGGATAATGA CCCTAAAGGT TTTAACCTGA
601 ---------+ ---------+ ---------+ ---------+ ---------+ 650
     K  E  N   G  Y  T  L   D  N  D   P  K  G   F  N  L  K

    AGTACCGTTA TGAACTCGAT GATAACTGGG GAGTAATAGG TTCGTTTGCT
651 ---------+ ---------+ ---------+ ---------+ ---------+ 700
      Y  R  Y   E  L  D   D  N  W  G   V  I  G   S  F  A

    TATACTCATC AGGGATATGA TTTCTTCTAT GGCAGTAATA AGTTTGGTCA
701 ---------+ ---------+ ---------+ ---------+ ---------+ 750
    Y  T  H  Q   G  Y  D   F  F  Y   G  S  N  K   F  G  H

    TGGTGATGTT GATTACTATT CAGTAACAAT GGGGCCATCT TTCCGCATCA
751 ---------+ ---------+ ---------+ ---------+ ---------+ 800
     G  D  V   D  Y  Y  S   V  T  M   G  P  S   F  R  I  N

    ACGAATATGT TAGCCTTTAT GGATTACTGG GGGCCGCTCA TGGAAAGGTT
801 ---------+ ---------+ ---------+ ---------+ ---------+ 850
      E  Y  V   S  L  Y   G  L  L  G   A  A  H   G  K  V
```

Figure 2 (continued)

```
      AAGGCATCTG TATTTGATGA ATCAATCAGT GCAAGTAAGA CGTCAATGGC
851   ----------+----------+----------+----------+----------+  900
      K  A  S  V   F  D  E   S  I  S   A  S  K  T   S  M  A

      ATACGGGGCA GGGGTGCAAT TCAACCCACT TCCAAATTTT GTCATTGACG
901   ----------+----------+----------+----------+----------+  950
       Y  G  A   G  V  Q  F   N  P  L   P  N  F   V  I  D  A

      CTTCATATGA ATACTCCAAA CTCGATAGCA TAAAAGTTGG CACCTGGATG
951   ----------+----------+----------+----------+----------+ 1000
        S  Y  E   Y  S  K   L  D  S  I   K  V  G   T  W  M

      CTTGGTGCAG GGTATCGATT CTAATCATCT CAGATAGTGA AAACCCACCT
1001  ----------+----------+----------+----------+----------+ 1050
      L  G  A  G   Y  R  F   *

      GAGTGAAGTG AACCCCATTT ATTGGACACT TTTCCTGGCG GTTGACATGG
1051  ----------+----------+----------+----------+----------+ 1100

      CCTGATTTCG GTACTGCACC GGACTCAGGC CGTTTAATTT TACTTTGATC
1101  ----------+----------+----------+----------+----------+ 1150

      CTTTCGTTGT TGTAGTAATG GATATACTCA TCCACCGCTT TTTTCAGTTG
1151  ----------+----------+----------+----------+----------+ 1200

      TTCTACATCT TCGTATTTTT CATTGTGCCA GCATTCAGTC TTCAGCAGAC
1201  ----------+----------+----------+----------+----------+ 1250

      CAAAAAAGTT TTCTATCACA GCATTATCCA GGCAGTTGCC CTTGCGCGAC
1251  ----------+----------+----------+----------+----------+ 1300

      ATACTTTGCT TTACTTCGCC AGACCCCAGC CTTTTCTTAT AGCTTGCCAT
1301  ----------+----------+----------+----------+----------+ 1350

      CTGATATTGC CAGCCCTGAT CCGAGTGAAG TACAGGTTCA TCGCCTGAGT
1351  ----------+----------+----------+----------+----------+ 1400

      TCAACTTCTG TAGCGCATCA TCAAGCATTT TATCAATCAG GTTCATTCCG
1401  ----------+----------+----------+----------+----------+ 1450

      GGATGCGTAT CCATCTGCCA GGCAACGACT TCGCTGTTAT ACAGATCCAG
1451  ----------+----------+----------+----------+----------+ 1500

      CACGGGTGAC AGATACAGCT TTTTACCCCT GACGTTGAAC TCGGTCACAT
1501  ----------+----------+----------+----------+----------+ 1550

      CGTTACCCAC TTCTGGTTAG GGGCTTCGGC AGTAAATTTT CGAGCAAGTA
1551  ----------+----------+----------+----------+----------+ 1600

      TATTAGGGAC CACTTTACCG TAGGCACCCT GATATGACTG ATATTTTTTA
1601  ----------+----------+----------+----------+----------+ 1650

      CGACGCAAGT TAGATGCAAG CTGCTGTTGC CGCATGAGTT TTCGTACGGT
1651  ----------+----------+----------+----------+----------+ 1700

      TTTATGGTTA AGACTCCCGC CCTCATTGCG TAGGGCCAGC GTTATTCTGC
1701  ----------+----------+----------+----------+----------+ 1750
```

Figure 2 (continued)

```
      GGTAACCATA GCGACCTTTA TGATGGTGAA ACAGGGTTTT TATTCTTTGT
1751  ----------+----------+----------+----------+----------+ 1800

      TTCTCATCCG CATAAGTCTC TTCACGACCA CTGGATTTTA CCTGCCAGTA
1801  ----------+----------+----------+----------+----------+ 1850

      GAAGGTGCTG CGCGGAAGAC CGGCCACGTA AAGCAAGGTC GCCAGTTTAT
1851  ----------+----------+----------+----------+----------+ 1900

      ACAGATGCCT TAATTCAGTG ATTATTCGCG TTTTTTCCGC TGCTTCTCTT
1901  ----------+----------+----------+----------+----------+ 1950

      ACAGGTGGTA TTCACTGAGT GCCACCGATA ATGCGCAGGC AAAGTCATTA
1951  ----------+----------+----------+----------+----------+ 2000

      ACGACCCCCG CCGCTCACCC TGAGCATGGT CGTTGATGGC TTTTATATTT
2001  ----------+----------+----------+----------+----------+ 2050

      TCCATAGAGC AGAGGATGAT TCTTTATGTC CCGAGTGAAC TGGGGTGAAC
2051  ----------+----------+----------+----------+----------+ 2100

      GGTTATCCCG GTTTGCCGCT GAATGGCAAC GGACGGGAAT ATCCCCTAAA
2101  ----------+----------+----------+----------+----------+ 2150

      GAGTGGTGTG AGAGAGAAGG TTATTCGTGG GGAACAGCGA AAGCGTATAT
2151  ----------+----------+----------+----------+----------+ 2200

      TTCGATAAAA GCAGCGAAAG
2201  ----------+----------+  2220
```

Figure 3

```
    ATCTTTTCCG TGGTATGACC AGAACATAAA GTTTTTGCTG CCCCAACGCC
1   ----------+----------+----------+----------+----------+   50

    GGTGTCAGGC GCATAACGCC TTCCAGCCGA TCTGCACTCA TCACGCCTGG
51  ----------+----------+----------+----------+----------+  100

    TTCCTAGTAG GTGAAATAAC TGCTGGGGAA GAAGGCTGAT GGGGTCATGT
101 ----------+----------+----------+----------+----------+  150

    TCTGATCAAG AATCAGCACG TTCGGCGCAA AAACGCTGGT TTGTTTGTTC
151 ----------+----------+----------+----------+----------+  200

    ACTTCGCTGG TCAGCGTCAG GGTCAGTTCG CCAATGTTTG CCGGGACGCT
201 ----------+----------+----------+----------+----------+  250

    GTACGCAGCA ACCGGACCAC TGATGCCGGG AACGTTCAGT TGTTGGCCGC
251 ----------+----------+----------+----------+----------+  300

    CGGTCGCCAG TTGGGTGGTC TGGGTTTTAG ATTGATCGAC CGGTGTCCAG
301 ----------+----------+----------+----------+----------+  350

    GTGAGTTGTT GCAGCGCAGC AGATGGAATG GCTGGCGCGT CGCTGGTGTT
351 ----------+----------+----------+----------+----------+  400

    TTGCGGTACG TAGTTAACAT CGGCAAGGCT AATTCCAGGC GCGCTTGCCA
401 ----------+----------+----------+----------+----------+  450

    GTAACCCTGC TGATAAACAG AGGACGATGA GACTTTTATT CATTTTCATT
451 ----------+----------+----------+----------+----------+  500

    GTTTTCACCT CAAAATCTGG AGCTCAGCGG TAGCCAGGCA ATAGCGCGCT
501 ----------+----------+----------+----------+----------+  550

    AAACCCGATA ATCAGAGGGG CTTTCGCCCC TTCAGATAAT GACAACCTGT
551 ----------+----------+----------+----------+----------+  600

    TTTTATGCCG GATGCGGCGT AAACGCCTTA TCCGGCCTAC ATTTGACAGC
601 ----------+----------+----------+----------+----------+  650

    CGTTGTAGGC CTGATAAGAC GCGCAAGCGT CGCATCAGGC GTTGGTTGCC
651 ----------+----------+----------+----------+----------+  700
```

Figure 3 (continued)

```
      GAATGCGGCG TAAACGCCTT ATCCGGCCCA GGTTTTGCTA TTACCACCAG
 701  ----------+----------+----------+----------+----------+  750

      ATTTCCATCT GGGCACCGAA GGTCCACTCG TCGCTGTCGC CACGACCGAA
 751  ----------+----------+----------+----------+----------+  800

      GCTGCCGCCG TTGAAATCAG CAGGAACGGC TTTGCCGAAG TTCGCGTTGT
 801  ----------+----------+----------+----------+----------+  850

      TATCAGCGTT ACCGGTGTAG TCGTAACCCC ATTTCTCATC CCACTTGGCG
 851  ----------+----------+----------+----------+----------+  900

      TAGGTTGCGA AGACACGAAT AGCCGGGCGT GACCAGATGC TGTCGCCAGC
 901  ----------+----------+----------+----------+----------+  950

      CTGCCATTGT TGTGCGAGGG TAATTTTGTA CTGATTGTTC TTGTCGCCGG
 951  ----------+----------+----------+----------+----------+ 1000

      TGCGCTGGGA TTCGACGTTG TCGTAGCCGA TTTCCATCAC GGTGCTCATG
1001  ----------+----------+----------+----------+----------+ 1050

      ATTGGCGTCC ACTTGTACAT CGGGCGAATA CCGACGGTCC ACCACTTGGT
1051  ----------+----------+----------+----------+----------+ 1100

      GCCGTTGTCG TTATCCCAGT TGATATCCTG GTACATACCC ACGTACATCA
1101  ----------+----------+----------+----------+----------+ 1150

      TGTCCCAGTT GTCGCCCATG GAGATCGCAC CGTGGTCGAG GATACGCAGC
1151  ----------+----------+----------+----------+----------+ 1200

      ATGTGACCGT TGTTGTTGAT ATTGTAGGCA AATTTTTCGT TATCAAATGC
1201  ----------+----------+----------+----------+----------+ 1250

      AACGCCAGAA CCCTGCGACA GCCCTTTACC CTGCGAGGTC ATCGAGTCAG
1251  ----------+----------+----------+----------+----------+ 1300

      TAGCGTACTG AACAACAAAC TTGTTAAAGC CCTTCAGGAC ACTCTGAGTA
1301  ----------+----------+----------+----------+----------+ 1350

      TGTTCAGCAG TGAATAACCA GCCGTCTTTC GATGCGCCAT CAACCAGACG
1351  ----------+----------+----------+----------+----------+ 1400

      ATAGTTATCA CGCAAGTTGG CACGACCGTA GTCGACACCC AGTTCTAATG
1401  ----------+----------+----------+----------+----------+ 1450
```

Figur 3 (continued)

```
                TGCCGCCCGG GTTGATTTCC ATCTGCGCTA AACGCACATC GAAAACGTCG
1451            ----------+ ----------+ ----------+ ----------+ ----------+ 1500

                TTCGCGGTTT CGTTGGTATA GTCATAAATA TTGTTGCTGG CGAAAGAGGA
1501            ----------+ ----------+ ----------+ ----------+ ----------+ 1550

                AGAACCACCA GCTTCAGAGG AGCGGGTTGC TGCCAGAGAG AGTTTACCGA
1551            ----------+ ----------+ ----------+ ----------+ ----------+ 1600

                AGCCAACATC GATGTTTTCC AGACCGGCAC CAGGACCAGA AATATCCCAG
1601            ----------+ ----------+ ----------+ ----------+ ----------+ 1650

                TAGTAGAAGT CGATCATATG AACGTCATGA CGTTGGTAGA AGCGCTTACC
1651            ----------+ ----------+ ----------+ ----------+ ----------+ 1700

                TGCCCAGATG GTGGAGCCTG GCAGCCATTC GATCAGGTTT TTACCCTGCA
1701            ----------+ ----------+ ----------+ ----------+ ----------+ 1750

                CGTTTGCTTC ACGGAAGGCC GGATCGGTAG CTTCCCAGTC ATTCTGTTGT
1751            ----------+ ----------+ ----------+ ----------+ ----------+ 1800

                GCGACGGAAT AGGCCACGTT AGTGTCGAAA TAGAAGCTCT TATCGCCCTC
1801            ----------+ ----------+ ----------+ ----------+ ----------+ 1850

                TTTCCACACT TCCTGACCCA ATTTTAATTC AGCATAAGTT TCACATTCGT
1851            ----------+ ----------+ ----------+ ----------+ ----------+ 1900

                TGCCAAGACG GTATTTACTT TGAGCACCGG TAGTCTGGAA ACACTGTTGT
1901            ----------+ ----------+ ----------+ ----------+ ----------+ 1950

                TCACCGCCGC TACCTGTCCA ACCAATACCG GAACGTGCAT AGCCGTGGAA
1951            ----------+ ----------+ ----------+ ----------+ ----------+ 2000

                ATCAACAGCC ATTGCCTGAG CAGACATTAC GCCCGCTGCG ACGGCAACCG
2001            ----------+ ----------+ ----------+ ----------+ ----------+ 2050

                CCAGAGGAAG TTTGCGCAGA GTAATCATCA TTCTATCTCC TGAGTCATTG
2051            ----------+ ----------+ ----------+ ----------+ ----------+ 2100

                CTTTTCTTTT TTCACATCAC CTGTGACAGG CTTTGTGTGT TTTGTGGGGT
2101            ----------+ ----------+ ----------+ ----------+ ----------+ 2150
```

Figure 3 (continued)

```
      GCTTAAACGC CCGGCTCCTT ATGCAGTCGA CGACATGCAG TGCCATCCTC
2151  ---------+ ---------+ ---------+ ---------+ ---------+  2200

      ACGGAACAGA TGGCAACGCT CTGGCGGCAG GCCGATAGCG AATGTGGCAC
2201  ---------+ ---------+ ---------+ ---------+ ---------+  2250

      CTTCTTCTAC CAACACCACG TCGTTCTGGC GGTACACCAG GTTTTGACGA
2251  ---------+ ---------+ ---------+ ---------+ ---------+  2300

      ATGGAAGGGA TCTGGATATG GATTTGAGTT TCGTTGCCGA GTTGCTCGAC
2301  ---------+ ---------+ ---------+ ---------+ ---------+  2350

      GACCTGAACT TCACCCTCAA GGATGACGTC AGCGATATCA CTCGGCAGTA
2351  ---------+ ---------+ ---------+ ---------+ ---------+  2400

      GATGTTCCGG GCGAATACCC AGCGACATAT TGGCTCCAAC CTGGACATCA
2401  ---------+ ---------+ ---------+ ---------+ ---------+  2450

      CGGCTTTCAA CTGGCAGCCA GACTTGCTGA CGATTTGGCA TCGGCAGCTC
2451  ---------+ ---------+ ---------+ ---------+ ---------+  2500

      CACCTGCACT TGATCGATTG CGGTGGCGGT CACTTTTACC GGCAGGAGTT
2501  ---------+ ---------+ ---------+ ---------+ ---------+  2550

      CATCTTTGGC GAACCGATAA ATCCGGCGAC AAAACGGTCT GCCGGATAGT
2551  ---------+ ---------+ ---------+ ---------+ ---------+  2600

      GGTACAGCTA GCGGTTTCCC AACCTGCGCC ACGCGACCGG CGTCCAGCAC
2601  ---------+ ---------+ ---------+ ---------+ ---------+  2650

      CACGATTTTG TCGGCCAGCG TCATCGCTTC GACCTGATCG TGGGTGACGT
2651  ---------+ ---------+ ---------+ ---------+ ---------+  2700

      AAATCATTGT GCGGCCCAGG CGTTTATGCA GACGGGAGAT TTCGATACGC
2701  ---------+ ---------+ ---------+ ---------+ ---------+  2750

      ATTTGCACAC GCAGTGCAGC ATCGAGGTTG GAGAGCGGTT CATCGAGCAA
2751  ---------+ ---------+ ---------+ ---------+ ---------+  2800

      AAATACGCTT GGCTCGGCCA CCAGCGTACG GCCAATCGCC ACACGCTGAC
2801  ---------+ ---------+ ---------+ ---------+ ---------+  2850
```

Figure 3 (continued)

```
      GCTGACCACC GGAGAGCGCT TTCGGTTTGC GATCCAGCAA ATGCGCCAGT
2851  ----------+----------+----------+----------+----------+  2900

      TGTAGCACTT CCGCCACCTG GTTAACGCGT TGGTTAATCA CCTCTTTTTT
2901  ----------+----------+----------+----------+----------+  2950

      TGCGCCAGCA GGTTTCAGGC CAAATGACAT GTTTTCTGCT ACTGACAGGT
2951  ----------+----------+----------+----------+----------+  3000

      GGGGATAGAG CGCGTAAGAC TGAAACACCA TACCAACGCC GCGTTCTGCT
3001  ----------+----------+----------+----------+----------+  3050

      GGCGGAGTGT CATTCATCCG TTTCTCACCG ATGAACAGGT CGCCGCTGGT
3051  ----------+----------+----------+----------+----------+  3100

      GATCGTCTCA AGCCCGGCAA TCATGCGCAG TAAAGTCGAT TTACCGCAGC
3101  ----------+----------+----------+----------+----------+  3150

      CAGACGGTCC GACAAACACC ACGAATTCAC CTTCATGGAT ATCGAGATTG
3151  ----------+----------+----------+----------+----------+  3200

      ATATCTTTCG ATACCACGAC CTCGCCCCAG GCTTTCGTTA CATTTTGCAG
3201  ----------+----------+----------+----------+----------+  3250

      CTGTACGCTC GCCATGCCCT TCTCCCTTTG TAACAACCTG TCATCGACAG
3251  ----------+----------+----------+----------+----------+  3300

      CAACATTCAT GATGGGCTGA CTATGCGTCA TCAGGAGATG GCTTAAATCC
3301  ----------+----------+----------+----------+----------+  3350

      TCCACCCCCT GGCTTTTTTA TGGGGGAGGA GGCGGGAGGA TGAGAACACG
3351  ----------+----------+----------+----------+----------+  3400

      GCTTCTGTGA ACTAAACCGA GGTCATGTAA GGAATTTCGT GATGTTGCTT
3401  ----------+----------+----------+----------+----------+  3450

      GCAAAAATCG TGGCGATTTT ATGTGCGCAT CTCCACATTA CCGCCAATTC
3451  ----------+----------+----------+----------+----------+  3500

      TGTAACAGAG ATCACACAAA GCGACGGTGG GGCGTAGGGG CAAGGAGGAT
3501  ----------+----------+----------+----------+----------+  3550
```

Figure 3 (continued)

```
                GGAAAGAGGT TGCCGTATAA AGAAACTAGA GTCCGTTTAG GTGTTTTCAC
3551            ---------+ ---------+ ---------+ ---------+ ---------+ 3600

                GAGCACTTCA CCAACAAGGA CCATAGATTA TGAAAATAAA AACAGGTGCA
3601            ---------+ ---------+ ---------+ ---------+ ---------+ 3650
                                              M   K   I   K   T   G   A

                CGCATCCTCG CATTATCCGC ATTAACGACG ATGATGTTTT CCGCCTCGGC
3651            ---------+ ---------+ ---------+ ---------+ ---------+ 3700
                R   I   L   A   L   S   A   L   T   T   M   M   F   S   A   S   A

                TCTCGCCAAA ATCGAAGAAG GTAAACTGGT AATCTGGATT AACGGCGATA
3701            ---------+ ---------+ ---------+ ---------+ ---------+ 3750
                 L   A   K   I   E   E   G   K   L   V   I   W   I   N   G   D   K

                AAGGCTATAA CGGTCTCGCT GAAGTCGGTA AGAAATTCGA GAAAGATACC
3751            ---------+ ---------+ ---------+ ---------+ ---------+ 3800
                  G   Y   N   G   L   A   E   V   G   K   K   F   E   K   D   T

                GGAATTAAAG TCACCGTTGA GCATCCGGAT AAACTGGAAG AGAAATTCCC
3801            ---------+ ---------+ ---------+ ---------+ ---------+ 3850
                G   I   K   V   T   V   E   H   P   D   K   L   E   E   K   F   P

                ACAGGTTGCG GCAACTGGCG ATGGCCCTGA CATTATCTTC TGGGCACACG
3851            ---------+ ---------+ ---------+ ---------+ ---------+ 3900
                 Q   V   A   A   T   G   D   G   P   D   I   I   F   W   A   H   D

                ACCGCTTTGG TGGCTACGCT CAATCTGGCC TGTTGGCTGA AATCACCCCG
3901            ---------+ ---------+ ---------+ ---------+ ---------+ 3950
                  R   F   G   G   Y   A   Q   S   G   L   L   A   E   I   T   P

                GACAAAGCGT TCCAGGACAA GCTGTATCCG TTTACCTGGG ATGCCGTACG
3951            ---------+ ---------+ ---------+ ---------+ ---------+ 4000
                D   K   A   F   Q   D   K   L   Y   P   F   T   W   D   A   V   R

                TTACAACGGC AAGCTGATTG CTTACCCGAT CGCTGTTGAA GCGTTATCGC
4001            ---------+ ---------+ ---------+ ---------+ ---------+ 4050
                 Y   N   G   K   L   I   A   Y   P   I   A   V   E   A   L   S   L

                TGATTTATAA CAAAGATCTG CTGCCGAACC CGCCAAAAAC CTGGGAAGAG
4051            ---------+ ---------+ ---------+ ---------+ ---------+ 4100
                  I   Y   N   K   D   L   L   P   N   P   P   K   T   W   E   E

                ATCCCGGCGC TGGATAAAGA ACTGAAAGCG AAAGGTAAGA GCGCGCTGAT
4101            ---------+ ---------+ ---------+ ---------+ ---------+ 4150
                I   P   A   L   D   K   E   L   K   A   K   G   K   S   A   L   M
```

Figure 3 (continued)

```
      GTTCAACCTG CAAGAACCGT ACTTCACCTG GCCGCTGATT GCTGCTGACG
4151  ----------+----------+----------+----------+----------+  4200
       F  N  L   Q  E  P  Y   F  T  W   P  L  I   A  A  D  G

      GGGGTTATGC GTTCAAGTAT GAAAACGGCA AGTACGACAT TAAAGACGTG
4201  ----------+----------+----------+----------+----------+  4250
        G  Y  A   F  K  Y   E  N  G  K   Y  D  I   K  D  V

      GGCGTGGATA ACGCTGGCGC GAAAGCGGGT CTGACCTTCC TGGTTGACCT
4251  ----------+----------+----------+----------+----------+  4300
       G  V  D  N   A  G  A   K  A  G   L  T  F  L   V  D  L

      GATTAAAAAC AAACACATGA ATGCAGACAC CGATTACTCC ATCGCAGAAG
4301  ----------+----------+----------+----------+----------+  4350
       I  K  N   K  H  M  N   A  D  T   D  Y  S   I  A  E  A

      CTGCCTTTAA TAAAGGCGAA ACAGCGATGA CCATCAACGG CCCGTGGGCA
4351  ----------+----------+----------+----------+----------+  4400
        A  F  N   K  G  E   T  A  M  T   I  N  G   P  W  A

      TGGTCCAACA TCGACACCAG CAAAGTGAAT TATGGTGTAA CGGTACTGCC
4401  ----------+----------+----------+----------+----------+  4450
       W  S  N  I   D  T  S   K  V  N   Y  G  V  T   V  L  P

      GACCTTCAAG GGTCAACCAT CCAAACCGTT CGTTGGCGTG CTGAGCGCAG
4451  ----------+----------+----------+----------+----------+  4500
       T  F  K   G  Q  P  S   K  P  F   V  G  V   L  S  A  G

      GTATTAACGC CGCCAGTCCG AACAAAGAGC TGGCGAAAGA GTTCCTCGAA
4501  ----------+----------+----------+----------+----------+  4550
        I  N  A   A  S  P   N  K  E  L   A  K  E   F  L  E

      AACTATCTGC TGACTGATGA AGGTCTGGAA GCGGTTAATA AAGACAAACC
4551  ----------+----------+----------+----------+----------+  4600
       N  Y  L  L   T  D  E   G  L  E   A  V  N  K   D  K  P

      GCTGGGTGCC GTAGCGCTGA AGTCTTACGA GGAAGAGTTG GCGAAAGATC
4601  ----------+----------+----------+----------+----------+  4650
       L  G  A   V  A  L  K   S  Y  E   E  E  L   A  K  D  P

      CACGTATTGC CGCCACCATG GAAAACGCCC AGAAAGGTGA AATCATGCCG
4651  ----------+----------+----------+----------+----------+  4700
        R  I  A   A  T  M   E  N  A  Q   K  G  E   I  M  P

      AACATCCCGC AGATGTCCGC TTTCTGGTAT GCCGTGCGTA CTGCGGTGAT
4701  ----------+----------+----------+----------+----------+  4750
       N  I  P  Q   M  S  A   F  W  Y   A  V  R  T   A  V  I
```

Figure 3 (continued)

```
      CAACGCCGCC AGCGGTCGTC AGACTGTCGA TGAAGCCCTG AAAGACGCGC
4751  ----------+----------+----------+----------+----------+ 4800
        N  A  A   S  G  R  Q   T  V  D   E  A  L   K  D  A  Q

      AGACTCGTAT CACCAAGTAA TGCTGTGAAA TGCCGGATGC GGCGTGAACG
4801  ----------+----------+----------+----------+----------+ 4850
        T  R  I   T  K  *

      CCTTGTCCGG CCTACAAAAC CGAAACGTAT GTAGGCCTGA TAAGACGCGT
4851  ----------+----------+----------+----------+----------+ 4900

      CAGCGTCGCA TCAGGCAGTT GTTGTCGGAT AAGGCGTGAA AGCCTTATCC
4901  ----------+----------+----------+----------+----------+ 4950

      GTCCTGGAAT GAGGAAGAAC CCCATGGATG TCATTAAAAA GAAACATTGG
4951  ----------+----------+----------+----------+----------+ 5000

      TGGCAAAGCG ACGCGCTGAA ATGGTCAGTG CTAGGTCTGC TCGGCCTGCT
5001  ----------+----------+----------+----------+----------+ 5050

      GGTGGGTTAC CTTGTTGTTT TAATGTACGC ACAAGGGGAA TACCTGTTCG
5051  ----------+----------+----------+----------+----------+ 5100

      CCATTACCAC GCTGATATTG AGTTCAGCGG GGCTGTATAT TTTCGCCAAT
5101  ----------+----------+----------+----------+----------+ 5150

      CGTAAAGCCT ACGCCTGGCG CTATGTTTAC CCGGGAATGG CTGGAATGGG
5151  ----------+----------+----------+----------+----------+ 5200

      ATTATTCGTC CTCTTCCCTC TGGTCTGCAC CATCGCCATT GCCTTCACCA
5201  ----------+----------+----------+----------+----------+ 5250

      ACTACAGCAG CACTAACCAG CTGACTTTTG AACGTGCGCA GGAAGTGTTG
5251  ----------+----------+----------+----------+----------+ 5300

      TTAGATCGCT CCTGGCAAGC AGGCAAAACC TATAACTTTG GTCTTTACCC
5301  ----------+----------+----------+----------+----------+ 5350

      GGCGGGCGAT GAGTGGCAAC TGGCGCTCAG CGACGGCGAA ACCGGCAAAA
5351  ----------+----------+----------+----------+----------+ 5400

      ATTACCTCTC CGACGCTTTT AAATTTGGCG GCGAGCAAAA ACTGCAACTG
5401  ----------+----------+----------+----------+----------+ 5450
```

Figure 3 (continued)

```
      AAAGAAACGA CCGCCCAGCC CGAAGGCGAA CGCGCGAATC TGCGCGTGAT
5451  ----------+ ---------+ ---------+ ---------+ ---------+  5500

      TACCCAGAAT CGTCAGGCGC TGAGTGACAT TACCGCCATT CTGCCGGATG
5501  ---------+ ---------+ ---------+ ---------+ ---------+  5550

      GCAACAAAGT GATGATGAGC TCCCTGCGCC AGTTTTCTGG CACGCAGCCG
5551  ---------+ ---------+ ---------+ ---------+ ---------+  5600

      CTCTACACAC TCGACGGTGA CGGCACGTTG ACGAATAATC AGAGCGGCGT
5601  ---------+ ---------+ ---------+ ---------+ ---------+  5650

      GAAATATCGT CCGAATAACC AAATTGGCTT TTACCAGTCC ATTACCGCCG
5651  ---------+ ---------+ ---------+ ---------+ ---------+  5700

      ACGGCAACTG GGGTGATGAA AAGCTAAGCC CCGGTTACAC CGTGACCACC
5701  ---------+ ---------+ ---------+ ---------+ ---------+  5750

      GGCTGGAAAA ACTTTACCCG CGTCTTTACC GACGAAGGCA TTCAGAAACC
5751  ---------+ ---------+ ---------+ ---------+ ---------+  5800

      GTTCCTCGCC ATTTTCGTCT GGACCGTGGT GTTCTCGCTG ATCACTGTCT
5801  ---------+ ---------+ ---------+ ---------+ ---------+  5850

      TTTTAACGGT GGCGGTCGGC ATGGTTCTGG CGTGTCTGGT GCAGTGGGAA
5851  ---------+ ---------+ ---------+ ---------+ ---------+  5900

      GCGTTGCGCG GCAAAGCGGT CTATCGCGTC CTGCTGATTC TGCCCTACGC
5901  ---------+ ---------+ ---------+ ---------+ ---------+  5950

      GGTGCCATCG TTCATTTCAA TCTTGATTTT CAAAGGGTTG TTTAACCAGA
5951  ---------+ ---------+ ---------+ ---------+ ---------+  6000

      GCTTCGGTGA AATCAACATG ATGTTGAGCG CGCTGTTTGG CGTGAAGCCC
6001  ---------+ ---------+ ---------+ ---------+ ---------+  6050

      GCCTGGTTCA GCGATCCGAC CACCGCCCGC ACGATGCTAA TTATCGTCAA
6051  ---------+ ---------+ ---------+ ---------+ ---------+  6100

      TACCTGGCTG GGTTATCCGT ACATGATGAT CCTCTGCATG GGCTTGCTGA
6101  ---------+ ---------+ ---------+ ---------+ ---------+  6150

      AAGCGATTCC GGACGATTTG TATGAAGCCT CAGCAATGGA TGGCGCAGGT
6151  ---------+ ---------+ ---------+ ---------+ ---------+  6200
```

Figure 3 (continued)

```
      CCGTTCCAGA ACTTCTTTAA GATTACGCTG CCGCTGCTGA TTAAACCGCT
6201  ----------+----------+----------+----------+----------+  6250

      GACGCCGCTG ATGATCGCCA GCTTCGCCTT TAACTTTAAC AACTTCGTGC
6251  ----------+----------+----------+----------+----------+  6300

      TGATTCAACT GTTAACCAAC GGCGGCCCGG ATCGTCTTGG CACGACCACG
6301  ----------+----------+----------+----------+----------+  6350

      CCAGCCGGTT ATACCGACCT GCTTGTTAAC TACACCTACC GCATCGCTTT
6351  ----------+----------+----------+----------+----------+  6400

      TGAAGGCGGC GGGGGTCAGG ACTTCGGTCT GGCGGCAGCA ATTGCCACGC
6401  ----------+----------+----------+----------+----------+  6450

      TGATCTTCCT GCTGGTGGGT GCGCTGGCGA TAGTGAACCT GAAAGCCACG
6451  ----------+----------+----------+----------+----------+  6500

      CGAATGAAGT TTGATTAAGG GAGATAACAA AAATGGCAAT GGTCC
6501  ----------+----------+----------+----------+-----  6545
```

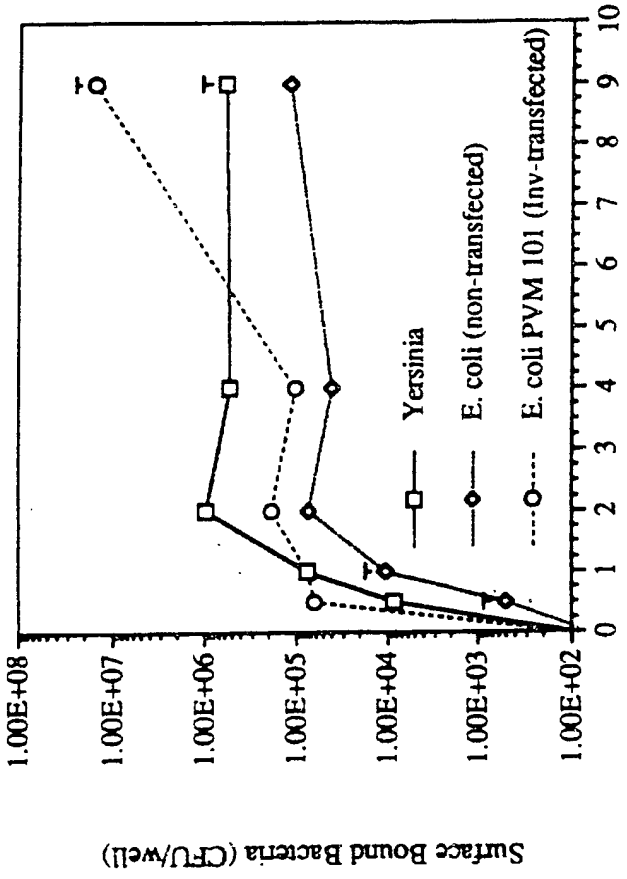
Surface Bound Bacteria (CFU/well)
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
0 1 2 3 4 5 6 7 8 9 10
Time (hrs)
Yersinia
E. coli (non-transfected)
E. coli PVM 101 (Inv-transfected)

Figure 4

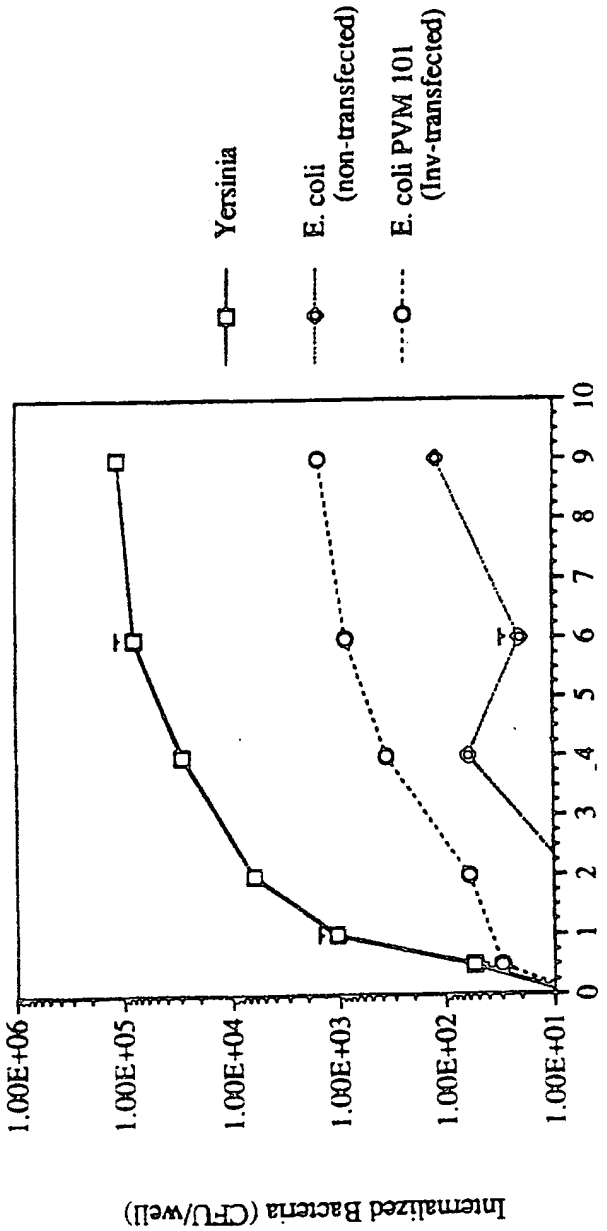
Yersinia
E. coli (non-transfected)
E. coli PVM 101 (Inv-transfected)
Internalized Bacteria (CFU/well)
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
0 1 2 3 4 5 6 7 8 9 10
Time (hrs)

Figure 5

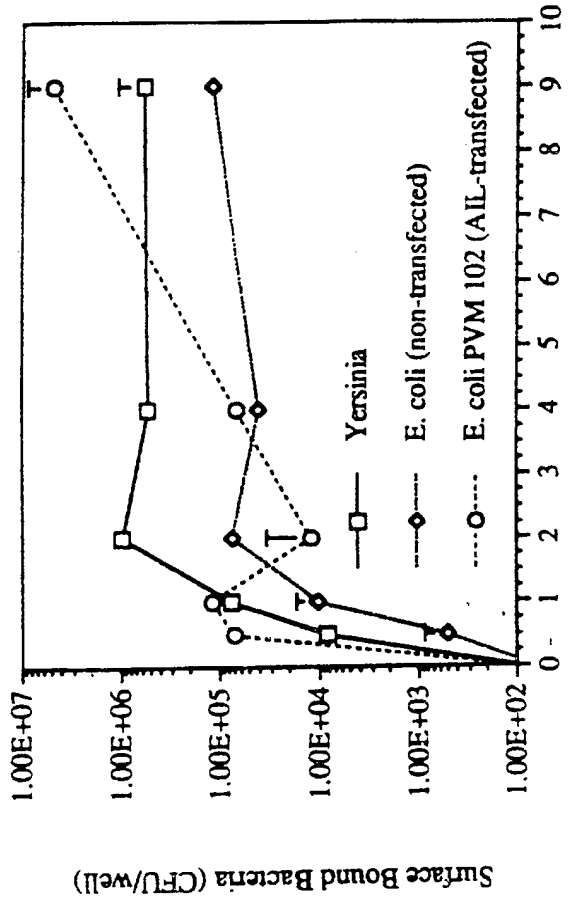
Surface Bound Bacteria (CFU/well)
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
0
1
2
3
4
5
6
7
8
9
10
Time (hrs)
Yersinia
E. coli (non-transfected)
E. coli PVM 102 (AIL-transfected)

Figure 6

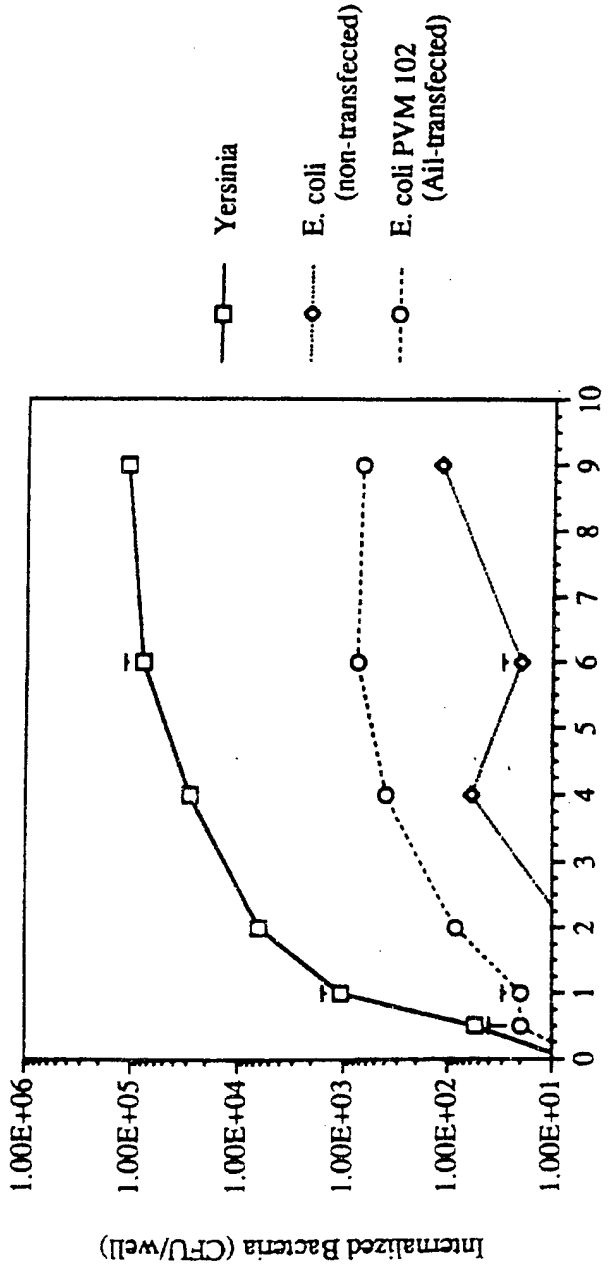
Yersinia
E. coli (non-transfected)
E. coli PVM 102 (Ail-transfected)
Internalized Bacteria (CFU/well)
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
0 1 2 3 4 5 6 7 8 9 10
Time (hrs)

Figure 7

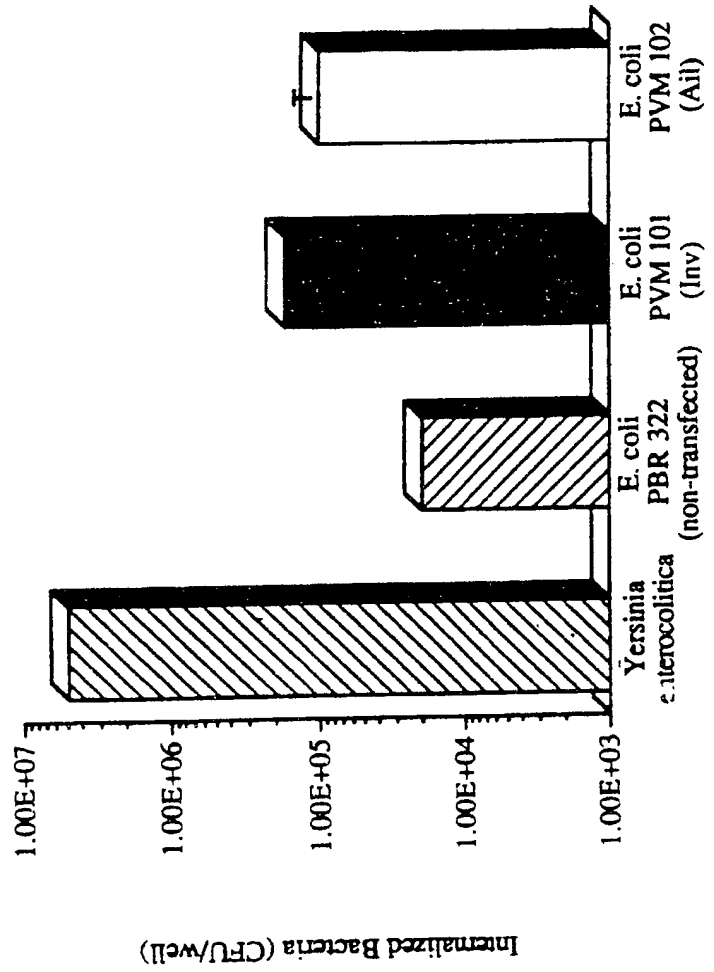
Internalized Bacteria (CFU/well)
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
Yersinia enterocolitica
E. coli PBR 322 (non-transfected)
E. coli PVM 101 (Inv)
E. coli PVM 102 (Ail)

Figure 8

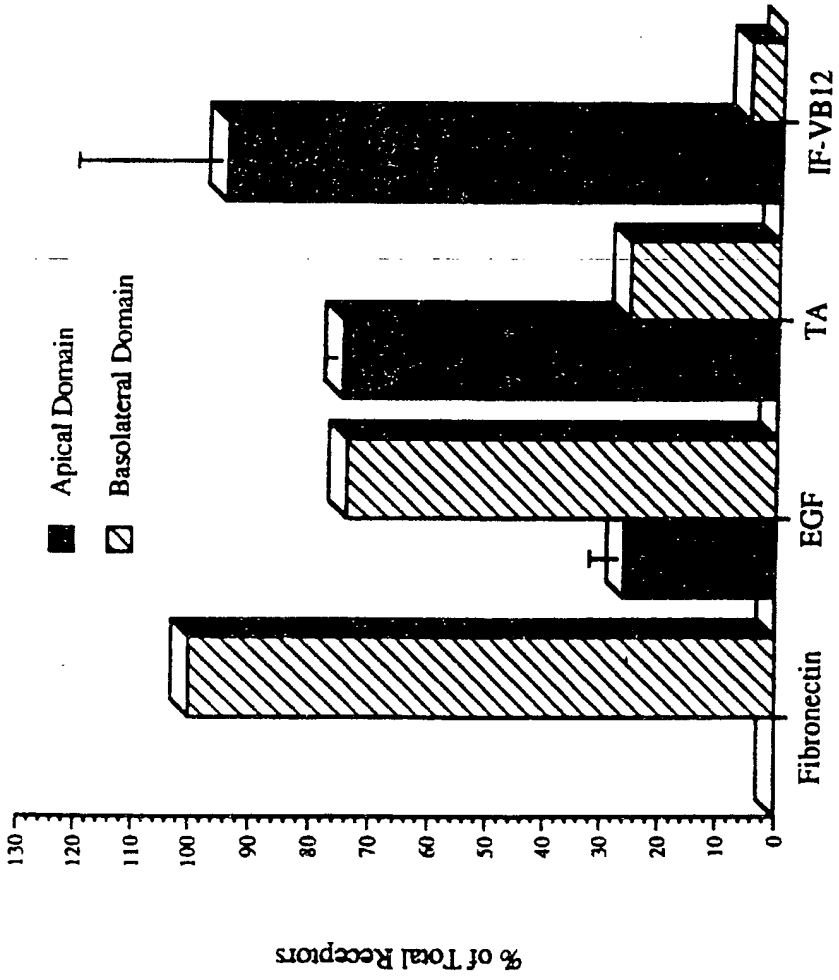
Apical Domain
Basolateral Domain
% of Total Receptors
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Fibronectin
EGF
TA
IF-VB12
Ligand

Figure 9

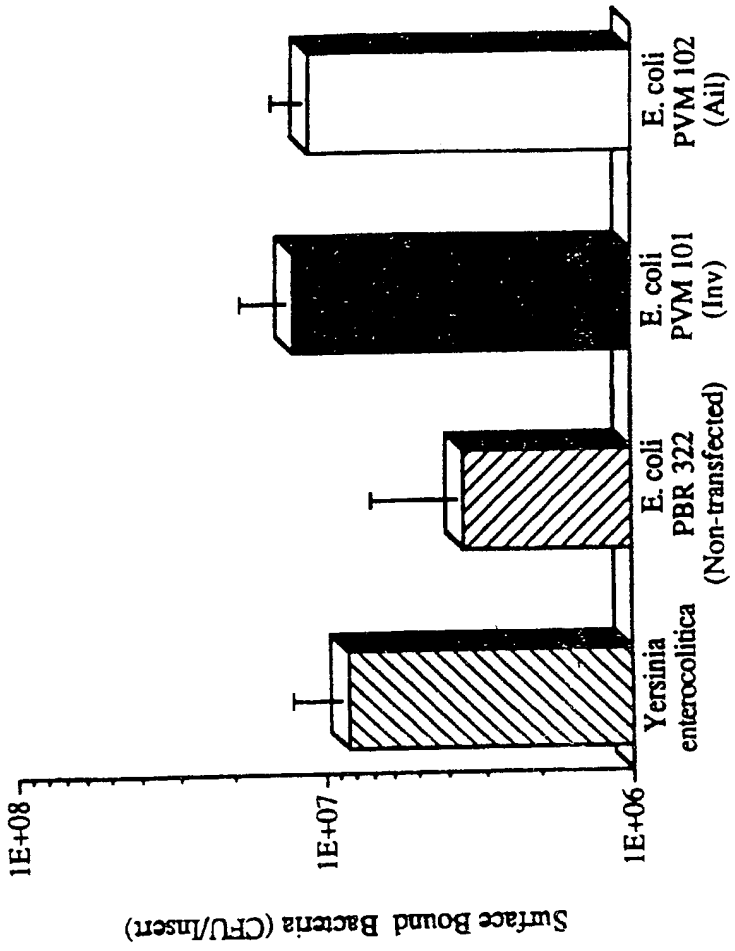
Surface Bound Bacteria (CFU/Insert)
1E+08
1E+07
1E+06
Yersinia enterocolitica
E. coli PBR 322 (Non-transfected)
E. coli PVM 101 (Inv)
E. coli PVM 102 (Ail)

Figure 10

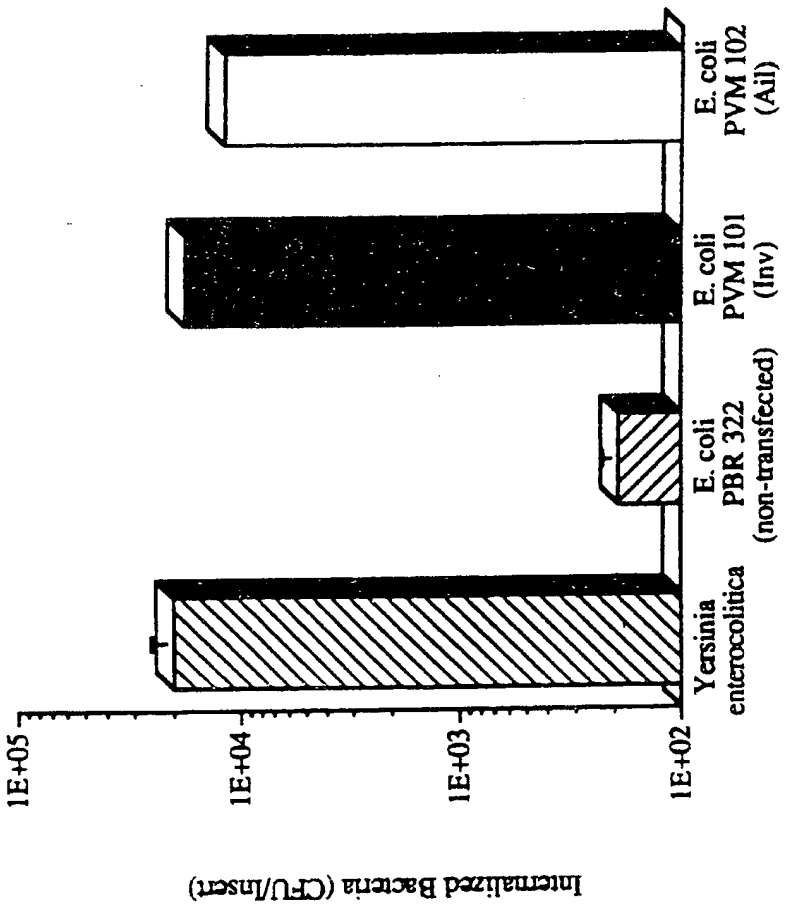
Internalized Bacteria (CFU/Insert)
1E+05
1E+04
1E+03
1E+02
Yersinia enterocolitica
E. coli PBR 322 (non-transfected)
E. coli PVM 101 (Inv)
E. coli PVM 102 (Ail)

Figure 11

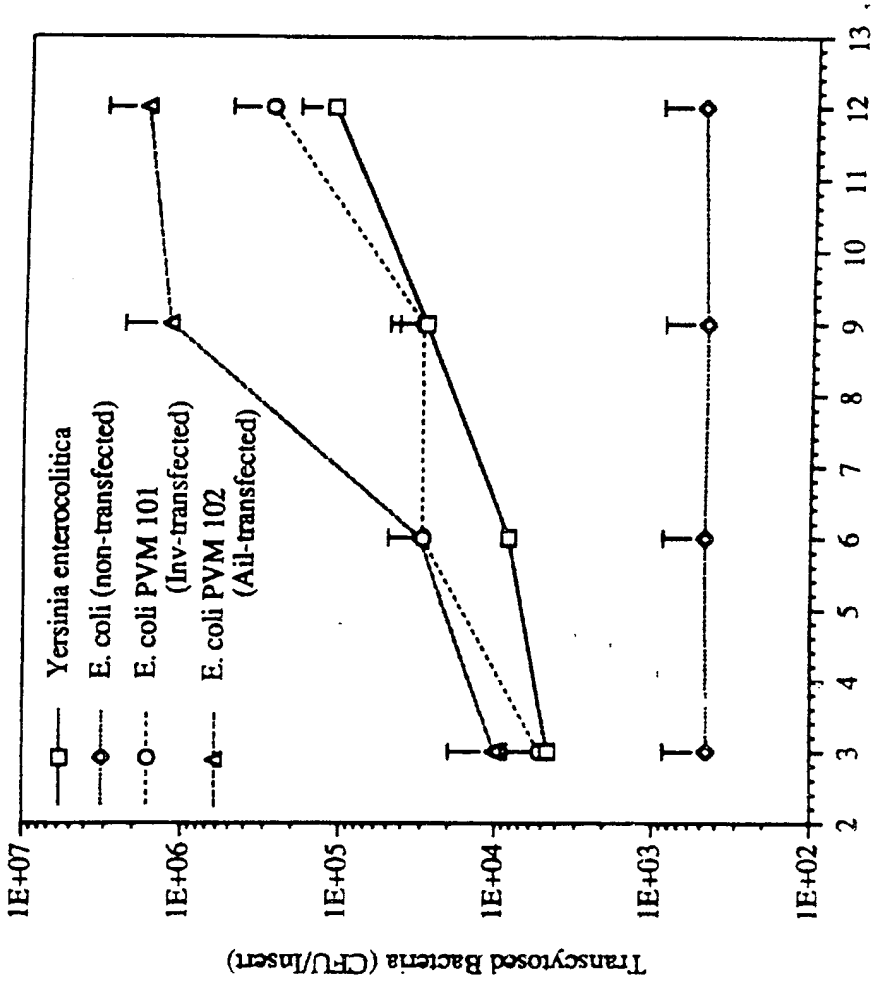
Yersinia enterocolitica
E. coli (non-transfected)
E. coli PVM 101 (Inv-transfected)
E. coli PVM 102 (Ail-transfected)
1E+07
1E+06
1E+05
1E+04
1E+03
1E+02
2
3
4
5
6
7
8
9
10
11
12
13
Time (hrs)
Transcytosed Bacteria (CFU/Insert)

Figure 12

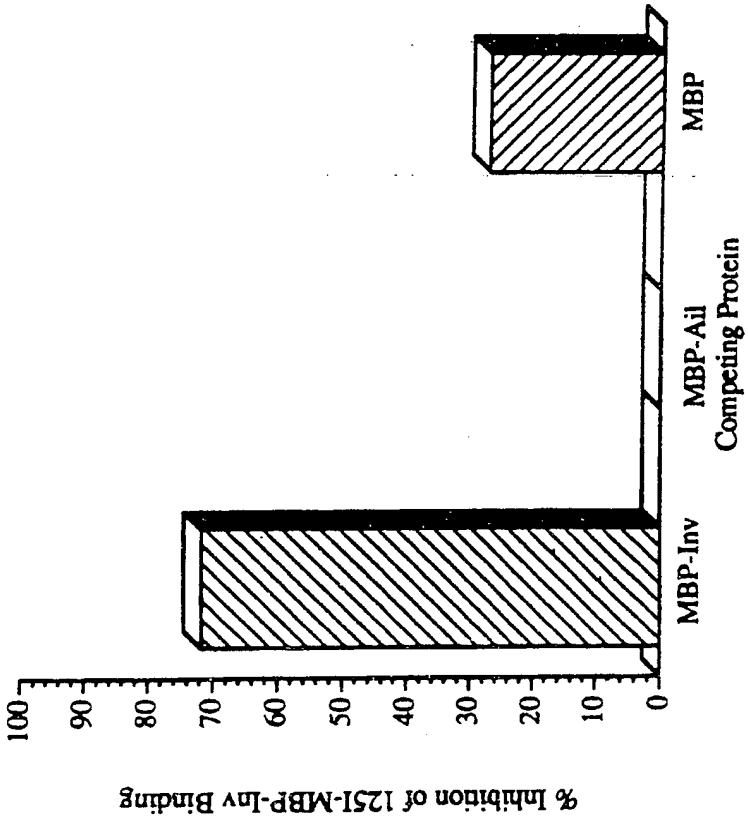
% Inhibition of 125I-MBP-Inv Binding
100
90
80
70
60
50
40
30
20
10
0
MBP-Inv
MBP-Ail
Competing Protein
MBP

Figure 13

Figure 14

MBP-INV(192) Fusion Protein

```
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1               5                  10                  15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
            20                  25                  30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
        35                  40                  45

Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
    50                  55                  60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65                  70                  75                  80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
                85                  90                  95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
            100                 105                 110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
        115                 120                 125

Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
    130                 135                 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145                 150                 155                 160

Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
                165                 170                 175

Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
            180                 185                 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
        195                 200                 205

Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
    210                 215                 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225                 230                 235                 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
                245                 250                 255

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
            260                 265                 270

Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
        275                 280                 285

Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
    290                 295                 300

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305                 310                 315                 320
```

Figure 14 (continued)

```
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
                325                 330                 335

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
            340                 345                 350

Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
        355                 360                 365

Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
    370                 375                 380

Glu Ala Leu Lys Asp Ala Gln Thr Arg Ile Thr Lys Val Pro Thr Leu
385                 390                 395                 400

Thr Gly Ile Leu Val Asn Gly Gln Asn Phe Ala Thr Asp Lys Gly Phe
                405                 410                 415

Pro Lys Thr Ile Phe Lys Asn Ala Thr Phe Gln Leu Gln Met Asp Asn
            420                 425                 430

Asp Val Ala Asn Asn Thr Gln Tyr Glu Trp Ser Ser Ser Phe Thr Pro
        435                 440                 445

Asn Val Ser Val Asn Asp Gln Gly Gln Val Thr Ile Thr Tyr Gln Thr
    450                 455                 460

Tyr Ser Glu Val Ala Val Thr Ala Lys Ser Lys Lys Phe Pro Ser Tyr
465                 470                 475                 480

Ser Val Ser Tyr Arg Phe Tyr Pro Asn Arg Trp Ile Tyr Asp Gly Gly
                485                 490                 495

Arg Ser Leu Val Ser Ser Leu Glu Ala Ser Arg Gln Cys Gln Gly Ser
            500                 505                 510

Asp Met Ser Ala Val Leu Glu Ser Ser Arg Ala Thr Asn Gly Thr Arg
        515                 520                 525

Ala Pro Asp Gly Thr Leu Trp Gly Glu Trp Gly Ser Leu Thr Ala Tyr
    530                 535                 540

Ser Ser Asp Trp Gln Ser Gly Glu Tyr Trp Val Lys Lys Thr Ser Thr
545                 550                 555                 560

Asp Phe Glu Thr Met Asn Met Asp Thr Gly Ala Leu Gln Pro Gly Pro
                565                 570                 575

Ala Tyr Leu Ala Phe Pro Leu Cys Ala Leu Ser Ile
            580                 585
```

Figure 15

MBP-AIL Fusion Protein

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45

Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65 70 75 80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125

Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160

Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
165 170 175

Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205

Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270

Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285

Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300

Figure 15 (continued)

```
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305                 310                 315                 320

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
                325                 330                 335

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
            340                 345                 350

Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala,
        355                 360                 365

Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
    370                 375                 380

Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser Val Pro Gly Arg
385                 390                 395                 400

Gly Ser Ile Glu Gly Arg Ala Ser Val Asn Val Tyr Ala Ala Ser Glu
                405                 410                 415

Ser Ser Ile Ser Ile Gly Tyr Ala Gln Ser His Val Lys Glu Asn Gly
            420                 425                 430

Tyr Thr Leu Asp Asn Asp Pro Lys Gly Phe Asn Leu Lys Tyr Arg Tyr
        435                 440                 445

Glu Leu Asp Asp Asn Trp Gly Val Ile Gly Ser Phe Ala Tyr Thr His
    450                 455                 460

Gln Gly Tyr Asp Phe Phe Tyr Gly Ser Asn Lys Phe Gly His Gly Asp
465                 470                 475                 480

Val Asp Tyr Tyr Ser Val Thr Met Gly Pro Ser Phe Arg Ile Asn Glu
                485                 490                 495

Tyr Val Ser Leu Tyr Gly Leu Leu Gly Ala Ala His Gly Lys Val Lys
            500                 505                 510

Ala Ser Val Phe Asp Glu Ser Ile Ser Ala Ser Lys Thr Ser Met Ala
        515                 520                 525

Tyr Gly Ala Gly Val Gln Phe Asn Pro Leu Pro Asn Phe Val Ile Asp
    530                 535                 540

Ala Ser Tyr Glu Tyr Ser Lys Leu Asp Ser Ile Lys Val Gly Thr Trp
545                 550                 555                 560

Met Leu Gly Ala Gly Tyr Arg Phe
                565
```

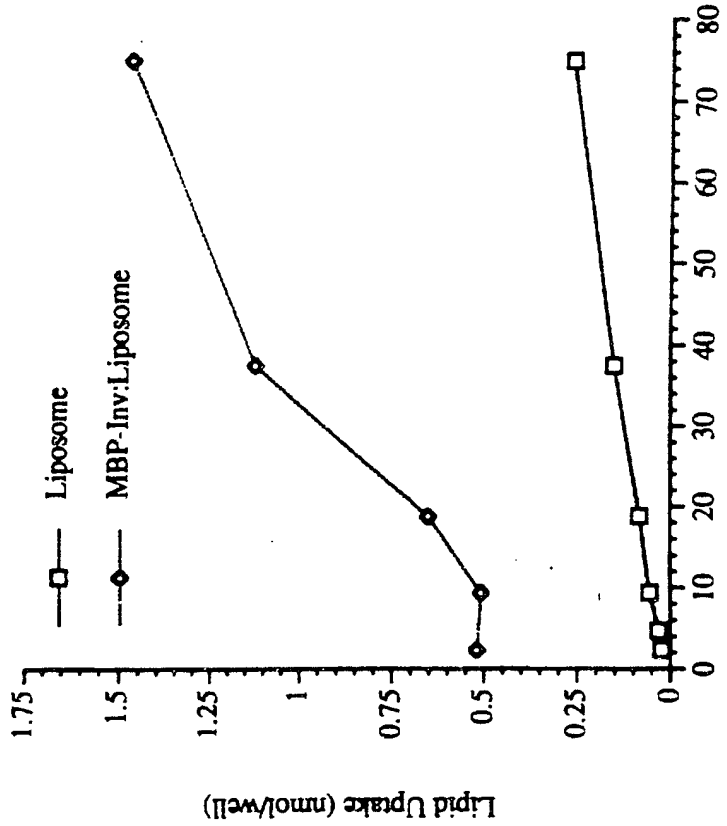
Liposome
MBP-Inv:Liposome
Lipid Uptake (nmol/well)
1.75
1.5
1.25
1
0.75
0.5
0.25
0
0
10
20
30
40
50
60
70
80
Lipid Added (uM)

Figure 16

INTERNATIONAL SEARCH REPORT

Interr nal Application No
PCT/US 95/13749

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 A61K9/127 A61K9/16 A61K9/51

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP,A,0 650 722 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 3 May 1995 see page 4, line 31 - line 36 see claims 1-6 ---	1-6, 8-16, 18-23
X	WO,A,94 18955 (ALZA CORPORATION) 1 September 1994 see the whole document see page 7, line 24 - page 8, line 5 ---	1-6, 8-16, 18-23 7,17
X	WO,A,92 17167 (BIOTECH AUSTRALIA PTY. LTD.) 15 October 1992 see page 24 - page 25; example 12 see claims 1-22 --- -/--	1-6, 8-16, 18-23

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
15 February 1996	23.02.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Benz, K

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 95/13749

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,90 04963 (DANBIOSYST UK LIMITED) 17 May 1990 see the whole document ---	1-6, 8-16, 18-23
Y	P. TYLE ET AL. 'TARGETED THERAPEUTIC SYSTEMS' 1990 , MARCEL DEKKER, INC. , NEW YORK (US) 169720 see page 163 - page 165, paragraph D -----	7,17

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13749

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 18-22 are directed to a method of treatment of the human/animal body the search has been carried out and based on the alleged effects of the composition.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Interna[illegible]al Application No
PCT/US 95/13749

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-650722	03-05-95	US-A-	5310654	10-05-94
		AU-B-	638456	01-07-93
		AU-B-	5562890	16-11-90
		EP-A-	0423317	24-04-91
		WO-A-	9012867	01-11-90
		US-A-	5239066	24-08-93
		US-A-	5338842	16-08-94
WO-A-9418955	01-09-94	AU-B-	6268894	14-09-94
		CA-A-	2151742	01-09-94
		EP-A-	0684814	06-12-95
WO-A-9217167	15-10-92	AU-B-	664365	16-11-95
		AU-B-	1558092	02-11-92
		CA-A-	2084194	03-10-92
		EP-A-	0531497	17-03-93
		NZ-A-	242220	27-04-94
WO-A-9004963	17-05-90	DE-D-	68909242	21-10-93
		DE-T-	68909242	17-02-94
		EP-A,B	0442949	28-08-91
		GB-A-	2243778	13-11-91
		JP-T-	4502910	28-05-92